FINAL REPORT

2009 ESTCP UXO Classification Study, San Luis Obispo, CA

ESTCP Project MR-200910

April 2012

Dean Keiswetter
SAIC

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
APR 2012		**00-00-2012 to 00-00-2012**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
2009 ESTCP UXO Classification Study, San Luis Obispo, CA	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **SAIC,1710 SAIC Drive,McLean,VA,22102**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**Same as Report (SAR)**	**161**	

Contents

Figures

Tables

Acronyms

11X	Depth corresponding to 11 times an object's diameter
CNG	California National Guard
DGM	Digital Geophysical Mapping
EE/CA	Engineering Evaluation/Cost Analysis
EMI	Electromagnetic Induction
ESTCP	Environmental Security Technology Certification Program
FUDS	Formerly Used Defense Site
GLRT	Generalized Likelihood Ratio Test
GPS	Global Positioning System
GSA	General Services Administration
HRR	Historical Records Review
IDA	Institute for Defense Analyses
MSEMS	Man Portable Simultaneous EMI and Magnetometer System
MRS	Munitions Response Site
MTADS	Multi-sensor Towed Array Detection System
N_{fa}	Number of False Alarms
NOSLN	No On Site Learning Necessary
NRL	Naval Research Laboratory
P_{class}	Probability of Correct Classification
QA	Quality Assurance
QC	Quality Control
ROC	Receiver Operating Characteristic
SAIC	Science Applications International Corporation
SCORR	Slope Corrected
SERDP	Strategic Environmental Research and Development Program
SI	Site Investigation
SLO	San Luis Obispo
SNR	Signal to Noise Ratio
TEMTADS	Time Domain EM Discrimination Array
TOI	Targets of Interest
UXO	Unexploded Ordnance

Acknowledgements

This work was funded by the Environmental Security Technology Certification Program Office, under projects MM-0910. The work was performed by a team of personnel from Science Applications International Corporation (SAIC).

We gratefully acknowledge Dr. Herb Nelson, Program Manager of ESTCP's Munitions Management Thrust Area, Dr. Anne Andrews, SERDP and ESTCP Deputy Director, and Dr. Jeff Marqusee, SERDP and ESTCP Executive Director, for their vision, support, and guidance throughout this demonstration.

The Principal Investigator for this project was Dr. Dean Keiswetter, SAIC. Key technical contributors from SAIC include Mr. Tom Furuya, Dr. Jim Kingdon, Mr. Jonathan Miller, and Dr. Tom Bell.

Executive Summary

ESTCP conducted a classification program at Camp San Luis Obispo in 2009 to demonstrate successful discrimination between UXO items and non-UXO item in situ. Three key components of their study included (i) the collection of high-quality geophysical data and principled selection of anomalous regions in those data, (ii) analysis of the selected anomalies using physics-based models to extract target parameters such as size, shape, and materials properties, and (iii) the use of those parameters to construct a prioritized dig list.

As part of the ESTCP study team, SAIC's objective was to discriminate targets of interest from native clutter at Camp San Luis Obispo, California, by characterizing and classifying anomalies identified in electromagnetic induction (EMI) and magnetic survey data. Both dynamic and cued (static) data were collected over the area. Misclassifying a target of interest as an item that can be left in the ground (viz., a false negative) was defined to be the primary failure. There were four main targets-of-interest at this site: 60-mm, 81-mm, 4.2-in mortars and 2.36-in rockets. During the excavation phase of the demonstration the three additional targets of interest (TOIs) were unearthed: 37mm projectile, 3inch stokes mortar, and 5inch rocket warhead.

The ESTCP Program Office coordinated data collection activities. This included all preparatory activities, arranging for a data collection by well-validated systems, selection of anomalies for analysis from each geophysical data set, and compilation of the individual sensor anomaly lists into a master list. Anomalies were selected from each geophysical data set using a target response-based threshold. In conjunction with the Advisory Group for the Demonstration, the ESTCP Program Office set the depth of interest for all items at this site as 45 cm below ground (depth measured to the center of the object). Validation digging was also coordinated by the Program Office

The data analysis was accomplished using UX-Analyze, an analysis framework within Oasis montaj™ that integrates quantitative analysis algorithms and custom-designed visualization schemes, to process MTADS magnetometer array, MTADS EM61 array, contractor EM61 cart, MSEMS EM61 array, and MSEMS magnetometer data. We analyzed TEMTADS and Metal Mapper data using in-house custom IDL inversion and processing routines.

The identities of approximately 150 recovered items were provided to SAIC for use as training data. These data were used as inputs to finalize classification algorithms and adjust thresholds. During the training process it was decided not to proceed with our analysis of the magnetic data due to the anticipated poor classification performance. The EM61 based data sets used the GLRT classification routine in UX-Analyze to classify and create the dig lists. The TEMTADS and Metal Mapper data classification used custom routines created in IDL. At the conclusion of training, SAIC submitted eight prioritized dig lists. Four dig lists were based on UX-Analyze processing of the survey data sets; including, MTADS EM61 array, contractor EM61 cart (with and without slope correction) and MSEMS EM61. The remaining four dig lists were submitted for the TEMTADS and Metal Mapper data. These lists were ordered from the item that is most confident not hazardous (Category 1) through the item that is most confident a munitions (Category 3). The anomalies in the transition zone between Category 1 and 3 for which we were

not able to make a decision were assigned Category 2. The anomalies for which we were not able to extract meaningful parameters (Category 4) were placed at the bottom of the list. Only Category 1 anomalies were recommended to be left in the ground. These inputs were scored by the Institute for Defense Analyses (IDA) with emphasis on the number of items that were correctly labeled non-hazardous while correctly labeling all munitions items.

In general, the four data sets collected in dynamic mode using the EM61 sensor produced similar and very good results. At our chosen operating threshold, the probability of correct classification was 1.0, 0.995 and 0.966 for the EM61 Array, EM61 Cart and EM61 MSEMS, respectively with clutter rejection rates of 49%, 57% and 67%. All the EM61 data sets used target size estimates and decay ratios to classify the anomalies. The individual polarizations were not of sufficient quality to use to calculate accurate shape information.

In contrast, the cued fixed array systems consisting of the TEMTADS and Metal Mapper produced polarizations that were accurate enough to discriminate between TOI and non-TOI on the basis of shape. At our chosen operating threshold, the probability of correct classification was 0.985 and 0.98 for the TEMTADS and Metal Mapper, respectively with clutter rejection rates of 70% and 74%. Classification for the TEMTADS and Metal Mapper datasets were primarily based on an algorithm which compared our derived polarizabilities with a library of known target signatures.

The false negatives can be divided into two main types. The first is comprised of low SNR targets. As the SNR decreases the accuracy of the inverted parameters also decreases. All the data underwent a thorough QC examination for data integrity but a few TOI, mainly from the Metal Mapper, inadvertently passed our checks. An accurate low signal cutoff was not determined for the Metal Mapper because we did not have background measurement data. The second category of false negatives is multiple targets. The majority of the false negatives for all systems fell under this category. Improvements here can proceed on two fronts. First, work is already underway to measure and compile a database of overlapping signatures with which we may further train our multi-target solver. In lieu of this, it will be necessary to derive a means of estimating from the data when there is a high likelihood of multiple targets present, and therefore, when we should not trust the results of a single dipole solver and classify the anomaly as Category 2.

1 INTRODUCTION

1.1 BACKGROUND

In 2003, the Defense Science Board observed: "The … problem is that instruments that can detect the buried UXOs also detect numerous scrap metal objects and other artifacts, which leads to an enormous amount of expensive digging. Typically 100 holes may be dug before a real UXO is unearthed! The Task Force assessment is that much of this wasteful digging can be eliminated by the use of more advanced technology instruments that exploit modern digital processing and advanced multi-mode sensors to achieve an improved level of discrimination of scrap from UXOs [1]."

Significant progress has been made in classification technology over the past several years. To date however, testing of these approaches has been primarily limited to test sites with only limited application at live sites. Acceptance of these classification technologies requires demonstration of system capabilities at real UXO sites under real world conditions. Any attempt to declare detected anomalies to be harmless and requiring no further investigation will require demonstration to regulators of not only individual technologies, but an entire decision making process.

The FY06 Defense Appropriation contained funding for the "Development of Advanced, Sophisticated, Discrimination Technologies for UXO Cleanup" in the Environmental Security Technology Certification Program (ESTCP). ESTCP responded by conducting a UXO Classification Study at the former Camp Sibert, AL [2]. The results of this first demonstration were very encouraging. Although conditions were favorable at the site, a single target-of-interest (4.2-in mortar) and benign topography and geology, all of the classification approaches demonstrated were able to correctly identify a sizable fraction of the anomalies as arising from non-hazardous items that could be safely left in the ground. Of particular note, the contractor EM61-MK2 cart survey with analysis using commercially-available methods correctly identified more than half the targets as non-hazardous.

To build upon the success of the first phase of this study, ESTCP sponsored a second study in 2008 at a site with more challenging topography and a wider mix of targets-of-interest (TOI). A range at the former Camp San Luis Obispo, CA was identified for this demonstration. This document describes SAIC's results of the demonstration at San Luis Obispo.

1.2 OBJECTIVE OF THE DEMONSTRATION

1.2.1 ESTCP UXO Discrimination Study

As outlined in the ESTCP UXO Discrimination Study Demonstration Plan [3], there were two primary objectives of this study:

1. Test and validate detection and classification capabilities of currently available and emerging technologies on real sites under operational conditions.
2. Investigate in cooperation with regulators and program managers how classification technologies can be implemented in cleanup operations.

Within each of these two overarching objectives, there were several sub-objectives.

Technical objectives of the Study

- Test and evaluate capabilities by demonstrating and evaluating individual sensor and classification technologies and processes that combine these technologies. Compare advanced methods to existing practices and validate the pilot technologies for the following:
  - Detection of UXOs
  - Identification of features that distinguish scrap and other clutter from UXO
  - Reduction of false alarms (items that could be safely left in the ground that are incorrectly classified as UXO) while maintaining Pds acceptable to all
  - Ability to identify sources of uncertainty in the classification process and to quantify their impact to support decision making, including issues such as impact of data quality due to how data is collected
  - Quantify the overall impact on risk arising from the ability to clear more land more quickly for the same investment.
  - Include the issues of a dig-no dig decision process and related QA/QC issues
- Understand the applicability and limitations of the pilot technologies in the context of project objectives, site characteristics, suspected ordnance contamination
- Collect high-quality, well documented data to support the next generation of signal processing research

1.2.2 Specific Objectives of the Demonstration

There were two specific demonstration objectives for SAIC's portion of the overall ESTCP demonstration. The first was to determine the discrimination capability, cost, and reliability of MTADS magnetometry data, MTADS EM61 MkII array data, Man portable Simultaneous Electromagnetic induction and Magnetometer System (MSEMS) EM61, MSEMS magnetometer, and standard cart EM61 MkII survey data using the commercially available analysis procedures found in UX-Analyze which was developed and demonstrated in MM-0210, *Feature based UXO Detection and Discrimination.* The second was to determine the discrimination capability, cost, and reliability of the NRL Cued Discrimination Array (aka TEMTADS) and Metal Mapper data using procedures based on those developed in ESTCP project MM-0601.

1.3 REGULATORY DRIVERS

ESTCP assembled an Advisory Group to address the regulatory, programmatic and stakeholder acceptance issues associated with the implementation of classification in the MR process. Additional details can be found in the ESTCP UXO Discrimination Study Demonstration Plan [3].

2 TECHNOLOGY

This demonstration consisted of data collection using a variety of geophysical sensor systems and analysis by several groups using a number of feature extraction and classification algorithms [4-7]. A brief description of the major components of the demonstration is provided below.

2.1 GEOPHYSICAL DATA COLLECTION

The two geophysical sensors employed in this demonstration were magnetometers and Electromagnetic Induction (EMI) sensors.

2.1.1 Magnetometer Sensor

Magnetometers sense the local perturbation to the earth's magnetic field due to nearby ferrous metal objects. Magnetometer survey data can be analyzed to obtain the location, depth and rough size of the ferrous target. For targets more than an object length or two away from the sensors, there is little shape information in the magnetometer response so these sensors are often more useful for detection than classification. The estimated depth from analysis of magnetometer data is very accurate in most cases and can be used to constrain the analysis of EMI data and improve the ultimate classification results.

A towed array of eight magnetometers with sensor location provided by cm-level GPS was used to acquire magnetometer data for this demonstration. Additional details can be found in the MTADS Magnetometer / MkII Demonstration Plan [4].

2.1.2 Electromagnetic Induction Sensors

A number of commercial and developmental EMI systems were demonstrated. SAIC analyzed data acquired using the Geonics EM61-MK2 sensor, the most widely used EMI sensor for UXO surveys, in two deployment techniques, one single-sensor cart configuration and an array of sensors. The single-sensor cart survey was performed by a NAEVA Geophysics using their existing commercial equipment. The array was a well-instrumented array with careful measurement of array position and orientation. Additional details can be found in the MTADS Magnetometer / MkII Demonstration Plan [4].

Two developmental EMI systems were used for cued interrogation. The Naval Research Lab MTADS Discrimination Array was used for cued interrogation of anomalies [5]. The Metal Mapper system was used in survey mode to identify anomalies and then returned in cued mode to classify the anomalies [6].

2.1.3 Dual-Mode Sensor

The final system demonstrated was a concurrent, magnetometer/EM61-MK2 dual mode sensor. This system operates by interleaving the magnetometer and EMI measurement periods to avoid interference between the sensors. This system was operated by NAEVA Geophysics with training and guidance provided by the SAIC which developed the MSEMS system [7].

2.2 DATA ANALYSIS

The data analysis was performed using UX-Analyze and custom IDL inversion and processing routines. We used UX-Analyze to process MTADS magnetomer array, MTADS EM61 array, contractor EM61 cart, MSEMS EM61 array, and MSEMS magnetometer data. The cued TEMTADS and Metal Mapper data were processed using in-house custom IDL inversion and processing routines.

2.2.1 UX-Analyze

UX-Analyze is a complete post-detection signal processing package developed under ESTCP funding. UX-Analyze allows users to systematically identify, extract, edit, and store data around individual anomalies. It provides efficient data structures and access for the analysis algorithms, stores the fitted parameters, and allows for multiple data types and surveys. The anomaly characterization algorithms in UX-Analyze assume a dipole source and derive the best set of induced dipole model parameters that account for the spatial variation of the signal as the sensor is moved over the object. The model parameters are target X,Y location and depth, three dipole response coefficients corresponding to the principle axes of the target (EMI only), and the three angles that describe the orientation of the target. The size of the target is estimated using empirical relationships between either the dipole moment for magnetic data or the sum of the targets' response coefficients. This module is the interface between Oasis montaj and the Demonstration analysis software (Figure 2-1).

Characterization routines for magnetic and EMI data have been integrated with the UX-Analyze framework. These 3-D routines include graphic displays and controls that allow the user to manually select and filter the input data for each anomaly (Figure 2-2).

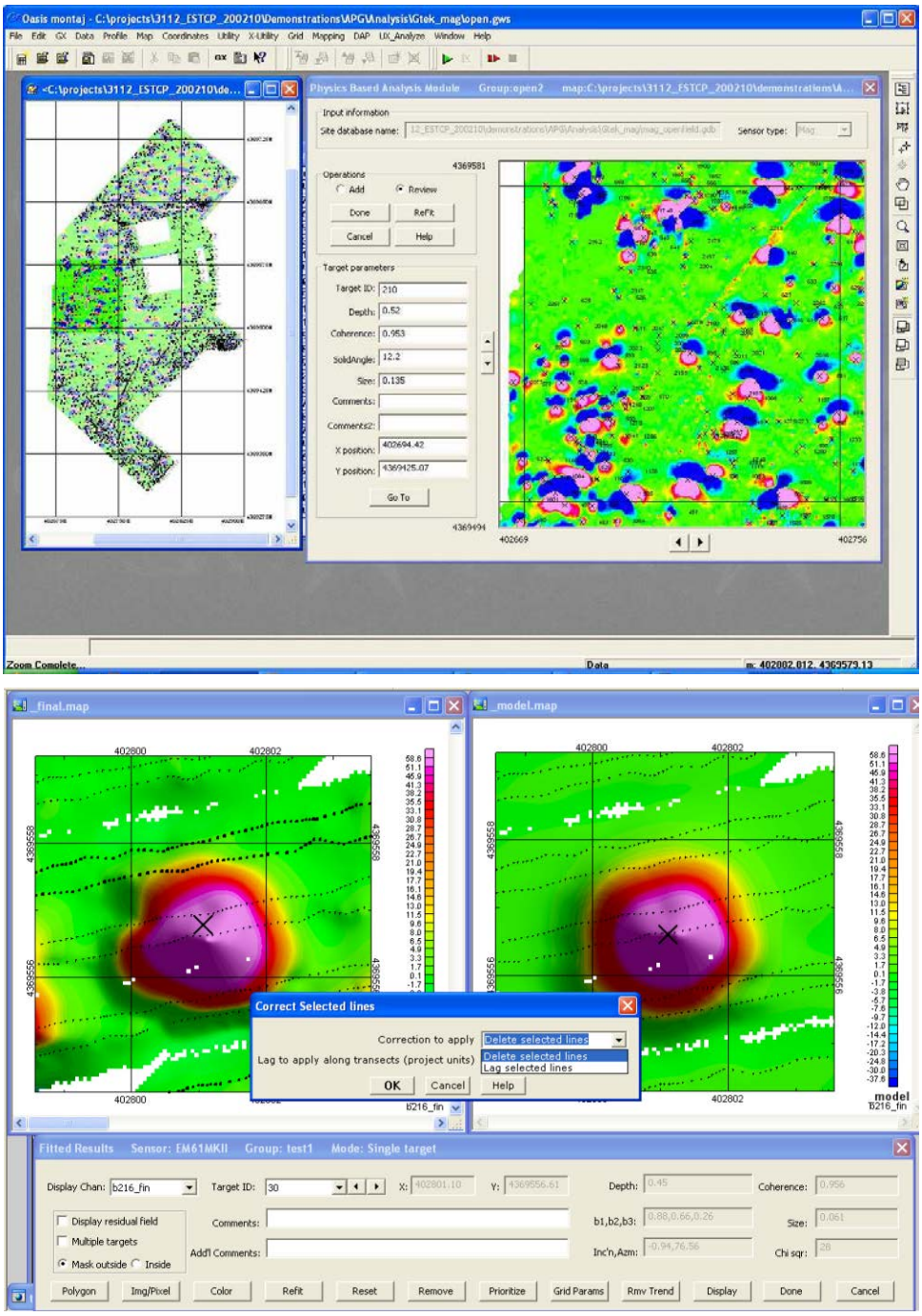


Figure 2-1. Snapshots of the computer monitor during analysis. The GUIs for searching, identifying, and reviewing anomalies is shown in the top image, while the modeling GUI is shown on the bottom. The control panel in the modeling GUI provides options for editing, gridding, polygoning, removing local trends, and displays results of the inversion. The large X symbols represent the location of the anomaly as measured (left map) and modeled (right).

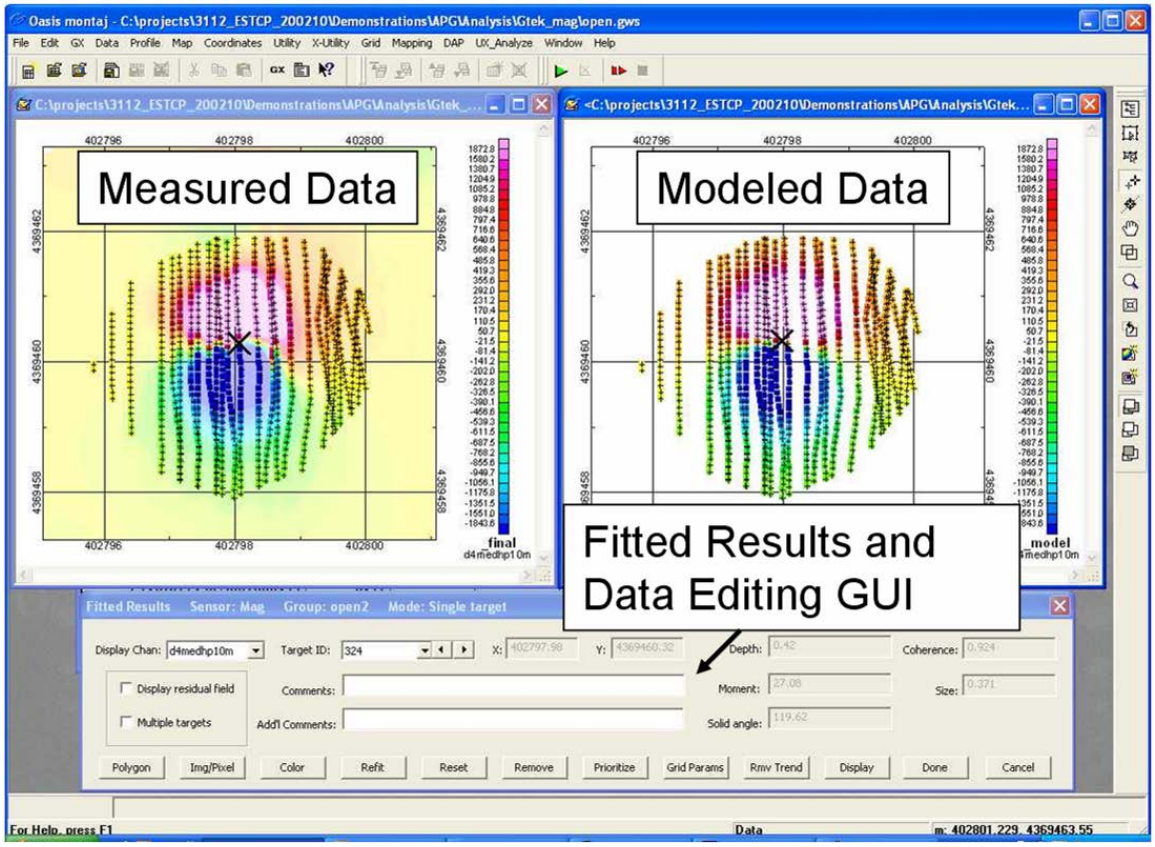


Figure 2-2. Screen snapshots showing the user interface during data inversion. The measured data is shown in the upper left map, the model parameters are displayed in the lower center window, and the forward model generated using the model parameters is shown in the upper right map.

UX-Analyze produces individualized anomaly reports, one for each anomaly, to document the decision process for each anomaly (Figure 2-3). In each plot, the measured data is graphically displayed next to the modeled data. The model parameters are listed in the middle of each page, and a profile extracted along the transect that passes closest to the dipoles location – as estimated by the inversion routine – is located at the bottom. The positions of individual measurements are superimposed on the maps. This layout was selected to provide insight into the confidence of the analysis and conclusions.

Essentially, the anomaly plots graphically provide an intuitive confidence measure. If the measured and modeled data are indistinguishable, the reviewer can have confidence that the estimated source parameters are approximately correct. If the two maps are do not resemble each other, however, it tells us that the source in question (i) cannot be represented well using a point dipole source, (ii) is not isolated, (iii) does not have sufficient signal-to-noise ratio, or (iv) was not properly sampled (spatially or temporally). In any case, if the two maps are dissimilar the inverted model parameters are most likely not correct.

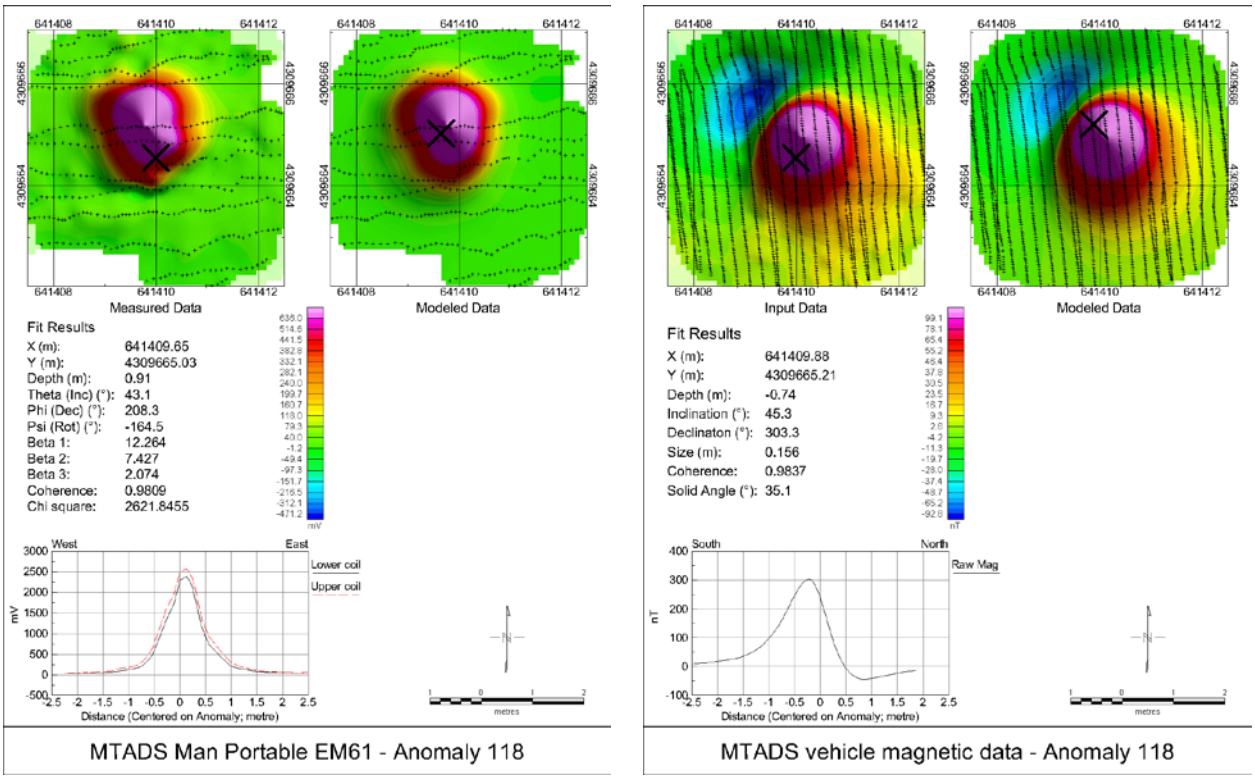


Figure 2-3. UX-Analyze generates a one-page summary for each anomaly. EMI data are shown in the left summary, and magnetic data on the right.

2.2.2 TEMTADS analysis

The TEMTADS EMI array consists of twenty-five individual EMI sensors arranged in a 5 x 5 array. The center-to-center distance is 40 cm yielding a 2m x 2m array. Transmitters are fired sequentially and recorded on all 25 receivers. Thus, there are 625 (25 x 25) individual transmit/receive pairs recorded. In Figure 2-4, we plot nine of the transmit/receive pairs resulting from excitation of a 40-mm projectile located under the center of the array. The decays plotted correspond to the signal received on the nine central sensors when that sensor transmits. In other words, the results of nine individual monostatic measurements are presented.

All 625 measurements are used for the inversion to recover target parameters. The inversion results for the decay data shown in Figure 2-4 are shown in Figure 2-5. As we expect for an object with axial symmetry such as a 40-mm projectile, we recover one large response coefficient and two equal, but smaller ones. These response coefficients will be the basis of the discrimination decisions in this demonstration. Derived βs for "Cylinder E" (3" x 12" steel cylinder) in the test field are shown for comparison in Figure 2-6.

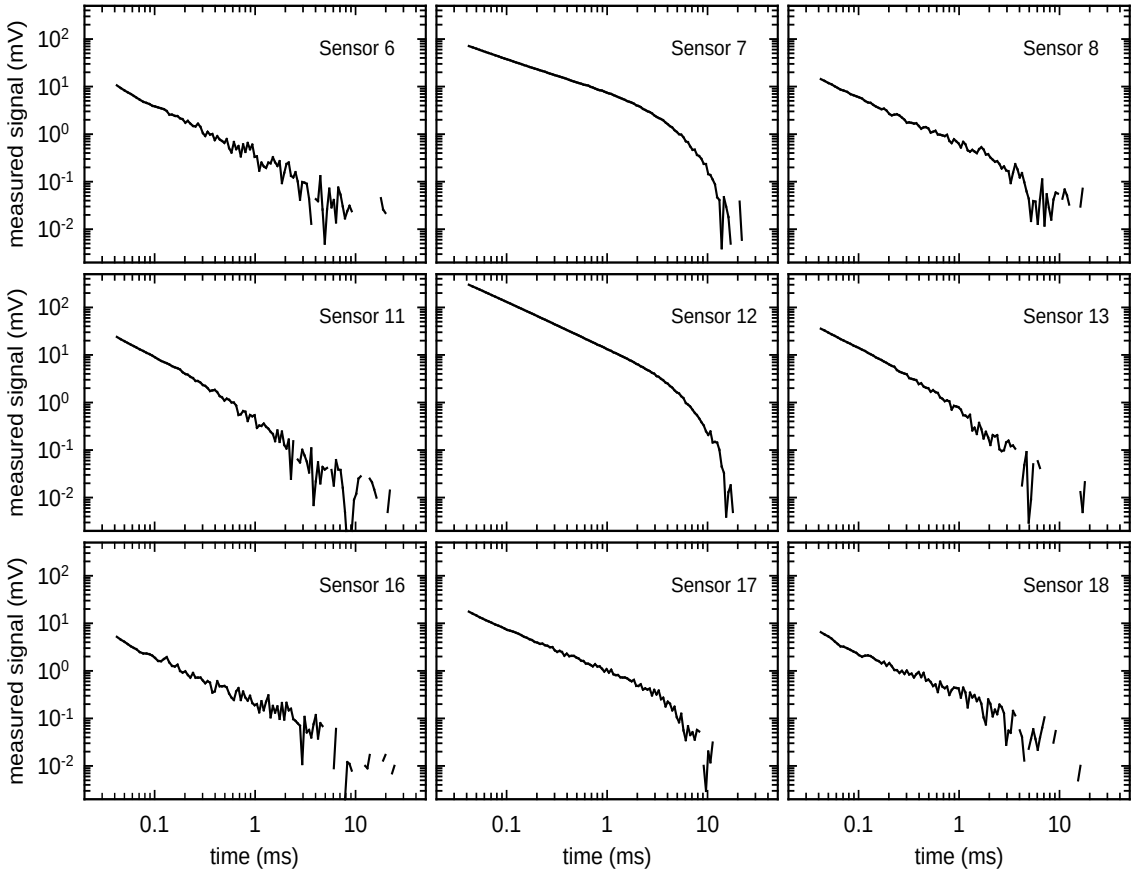


Figure 2-4. The response of nine of the individual sensors to a 40-mm projectile located under the center of the array.

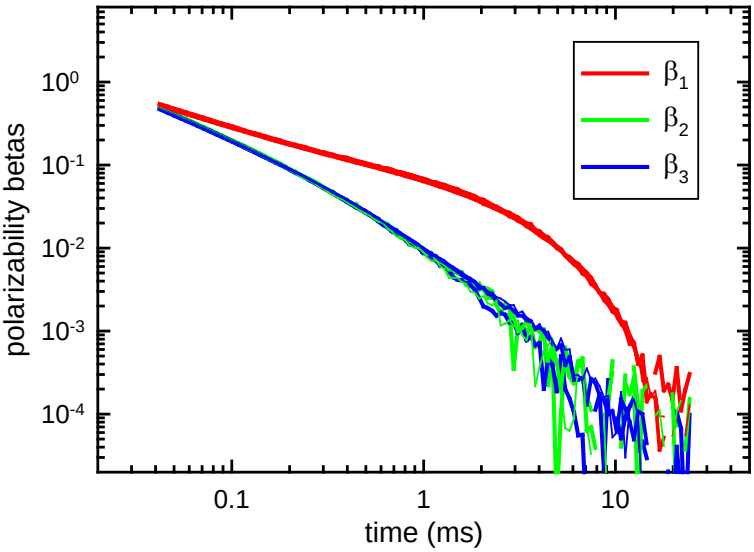


Figure 2-5. Derived response coefficients for a 40-mm projectile using the measurements of which the decays shown in Figure 2-4 are a subset.

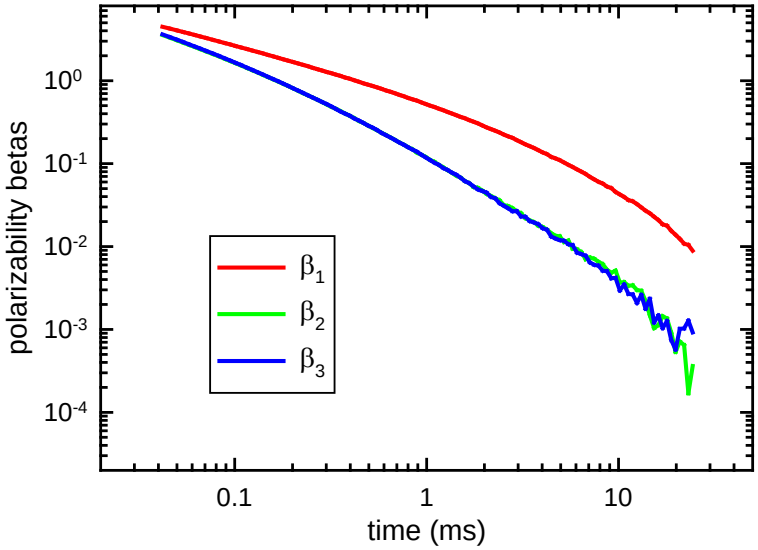


Figure 2-6. Derived response coefficients from a cued measurement over "Cylinder E" in the test field.

2.2.3 Metal Mapper Analysis

The Metal Mapper consists of three, one-meter square, orthogonal, high power transmitter loops with seven tri-axial receiver cubes mounted within the horizontal transmitter coil. The receiver coils are wire wrapped on 10cm x10cm cubes. The system was deployed in dynamic mapping mode where only the horizontal transmitter was used to map the site and detect targets. Once the target were identified the Metal Mapper was deployed in static mode where the array is placed stationary above that target and all three transmitter coils are fired sequentially and a longer recording window is employed. Thus in static mode, there are 63 (3 x 7 x 3) individual transmit/receive pairs recorded.

All 63 measurements are used for the inversion to recover target parameters. The inversions routines were essentially the same as those for the TEMTADS with adjustments being made for the different size and configuration of the transmitter and receiver coils.

2.2.4 Classification

Several types of classification processing were evaluated. These included a Generalized Likelihood Ratio Test (GLRT), and a formal rule-based classification scheme.

2.3 ADVANTAGES AND LIMITATIONS OF THE TECHNOLOGY

This Demonstration technology uses spatially referenced geophysical data to estimate model parameters. This has an inherent advantage over analysis methods based on non-quantitative schemes. Non-quantitative analysis methods frequently base the declaration on anomaly amplitude, half width, spatial footprint, or overall 'look'. These characterization methods are, however, sensitive to the targets' orientation and depth of burial. The methodology demonstrated here separates the measured signatures into that which is inherent to the target, and that which is related to the geometry of the problem, such as distance to sensor and orientation.

Known limitations to the data analysis approach adopted here result from (i) non-unique inversion results, and (ii) overlapping, or non distinct, signatures in feature space. The former limitation, one in which multiple sets of model parameters explain the vast majority of the observed data, is well known. The second, while perhaps not as widely appreciated, is equally problematic. Inverting EMI data using our dipole models, results in three eigenvalues of the magnetic polarizability tensor, each corresponds to a principal axis of object. Discrimination is possible only to the degree that the derived eigenvalues are different for different objects and stable for similar classes of objects. In other words, even with ideal data, the estimated burial depth, apparent size, and shape features may not separate UXO and clutter signatures into distinct, non-overlapping classes. This is because the anomaly features derived from EMI and magnetic data are not unique to UXO. Clutter items that have similar shapes and burial attributes to munitions can have geophysical signatures that are indistinguishable from UXO signatures and, as such, will have similar eigenvalues and therefore likely be classified as munitions. Examples include items such as pipes, post sections and axial symmetrical fragments.

3 PERFORMANCE OBJECTIVES

The performance objectives for this demonstration are summarized in Table 3-1. Since this was a classification demonstration, the performance objectives focused on target analysis and classification; we assumed that the anomalies from all targets of interest have been detected and included on the target list the analysis demonstrators will work from.

The first three objectives refer to the classification part of the demonstration with the first two referring to the best results from each approach in a retrospective analysis and the third addressing how well the correct threshold is specified in advance. The final two objectives refer to the feature extraction part of the demonstration.

Table 3-1 Performance Objectives for EM61 MK2 Array

Performance Objective	Metric	Data Required	Success Criteria	Results
Quantitative Performance Objectives				
Maximize correct classification of munitions	Number of targets-of-interest retained.	• Prioritized anomaly lists • Scoring reports from IDA	Approach correctly classifies all targets-of-interest	206 of 206 (100%) correctly classified as TOI
Maximize correct classification of non-munitions	Number of false alarms eliminated.	• Prioritized anomaly lists • Scoring reports from IDA	Reduction of false alarms by > 30% while retaining all targets of interest	565 of 1076 (52%) correctly classified as non-munitions
Specification of no-dig threshold	P_{class} and N_{fa} at demonstrator operating point.	• Demonstrator - specified threshold • Scoring reports from IDA	Threshold specified by the demonstrator to achieve criteria above	P_{class} = 1.00 N_{fa} = 551
Minimize number of anomalies that cannot be analyzed	Number of anomalies that must be classified as "Unable to Analyze."	• Demonstrator target parameters	Reliable target parameters can be estimated for > 90% of anomalies on each sensor's detection list.	11 of 1282 (0.9%) classified as "Unable to analyze"
Correct estimation of target parameters	Accuracy of estimated target parameters.	• Demonstrator target parameters • Results of intrusive investigation	βs : COV < 0.20 X, Y < 15 cm (1σ) Z < 10 cm (1σ) Size: COV < 0.20	β1,2,3 COV=0.83,0.76,0.77 XY mean = 24cm (14cm σ) Z mean= -19cm (20cm σ) Size COV=0.195

Table 3-2 Performance Objectives for EM61 MK2 Cart – Slope corrected

Performance Objective	Metric	Data Required	Success Criteria	Results
Quantitative Performance Objectives				
Maximize correct classification of munitions	Number of targets-of-interest retained.	• Prioritized anomaly lists • Scoring reports from IDA	Approach correctly classifies all targets-of-interest	211 of 212 (99.5%) correctly classified as TOI
Maximize correct classification of non-munitions	Number of false alarms eliminated.	• Prioritized anomaly lists • Scoring reports from IDA	Reduction of false alarms by > 30% while retaining all targets of interest	592 of 1074 (55%) correctly classified as non-munitions
Specification of no-dig threshold	P_{class} and N_{fa} at demonstrator operating point.	• Demonstrator - specified threshold • Scoring reports from IDA	Threshold specified by the demonstrator to achieve criteria above	P_{class} = 0.995 N_{fa} = 461
Minimize number of anomalies that cannot be analyzed	Number of anomalies that must be classified as "Unable to Analyze."	• Demonstrator target parameters	Reliable target parameters can be estimated for > 90% of anomalies on each sensor's detection list.	1 of 1286 (0.08%) classified as "Unable to analyze"
Correct estimation of target parameters	Accuracy of estimated target parameters.	• Demonstrator target parameters • Results of intrusive investigation	βs : COV < 0.20 X, Y < 15 cm (1σ) Z < 10 cm (1σ) Size: COV < 0.20	β1,2,3 COV=0.48,0.47,0.50 XY mean = 22cm (13cm σ) Z mean= -13cm (10cm σ) Size COV=0.13

Table 3-2 Performance Objectives for EM61 MK2 Cart – Not Slope corrected

Performance Objective	Metric	Data Required	Success Criteria	Results
Quantitative Performance Objectives				
Maximize correct classification of munitions	Number of targets-of-interest retained.	• Prioritized anomaly lists • Scoring reports from IDA	Approach correctly classifies all targets-of-interest	211 of 212 (99.5%) correctly classified as TOI
Maximize correct classification of non-munitions	Number of false alarms eliminated.	• Prioritized anomaly lists • Scoring reports from IDA	Reduction of false alarms by > 30% while retaining all targets of interest	589 of 1074 (55%) correctly classified as non-munitions
Specification of no-dig threshold	P_{class} and N_{fa} at demonstrator operating point.	• Demonstrator - specified threshold • Scoring reports from IDA	Threshold specified by the demonstrator to achieve criteria above	P_{class} = 0.995 N_{fa} = 459
Minimize number of anomalies that cannot be analyzed	Number of anomalies that must be classified as "Unable to Analyze."	• Demonstrator target parameters	Reliable target parameters can be estimated for > 90% of anomalies on each sensor's detection list.	2 of 1286 (0.16%) classified as "Unable to analyze"
Correct estimation of target parameters	Accuracy of estimated target parameters.	• Demonstrator target parameters • Results of intrusive investigation	βs : COV < 0.20 X, Y < 15 cm (1σ) Z < 10 cm (1σ) Size: COV < 0.20	β1,2,3 COV=0.48,0.48,0.49 XY mean = 31cm (17cm σ) Z mean= -13cm (11cm σ) Size COV=0.14

Table 3-3 Performance Objectives for EM61 MSEMS

Performance Objective	Metric	Data Required	Success Criteria	Results
Quantitative Performance Objectives				
Maximize correct classification of munitions	Number of targets-of-interest retained.	• Prioritized anomaly lists • Scoring reports from IDA	Approach correctly classifies all targets-of-interest	198 of 205 (96.6%) correctly classified as TOI
Maximize correct classification of non-munitions	Number of false alarms eliminated.	• Prioritized anomaly lists • Scoring reports from IDA	Reduction of false alarms by > 30% while retaining all targets of interest	48 of 1099 (4%) correctly classified as non-munitions
Specification of no-dig threshold	P_{class} and N_{fa} at demonstrator operating point.	• Demonstrator - specified threshold • Scoring reports from IDA	Threshold specified by the demonstrator to achieve criteria above	P_{class} = 0.966 N_{fa} = 360
Minimize number of anomalies that cannot be analyzed	Number of anomalies that must be classified as "Unable to Analyze."	• Demonstrator target parameters	Reliable target parameters can be estimated for > 90% of anomalies on each sensor's detection list.	0 of 1304 (0.0%) classified as "Unable to analyze"
Correct estimation of target parameters	Accuracy of estimated target parameters.	• Demonstrator target parameters • Results of intrusive investigation	βs : COV < 0.20 X, Y < 15 cm (1σ) Z < 10 cm (1σ) Size: COV < 0.20	β1,2,3 COV=0.52,0.50,0.65 XY mean = 21cm (12cm σ) Z mean= -6cm (13cm σ) Size COV=0.17

Table 3-4 Performance Objectives for TEMTADS – 2 criteria

Performance Objective	Metric	Data Required	Success Criteria	Results
Quantitative Performance Objectives				
Maximize correct classification of munitions	Number of targets-of-interest retained.	• Prioritized anomaly lists • Scoring reports from IDA	Approach correctly classifies all targets-of-interest	203 of 206 (98.5%) correctly classified as TOI
Maximize correct classification of non-munitions	Number of false alarms eliminated.	• Prioritized anomaly lists • Scoring reports from IDA	Reduction of false alarms by > 30% while retaining all targets of interest	71 of 1076 (6.6%) correctly classified as non-munitions
Specification of no-dig threshold	P_{class} and N_{fa} at demonstrator operating point.	• Demonstrator - specified threshold • Scoring reports from IDA	Threshold specified by the demonstrator to achieve criteria above	P_{class} = 0.985 N_{fa} = 324
Minimize number of anomalies that cannot be analyzed	Number of anomalies that must be classified as "Unable to Analyze."	• Demonstrator target parameters	Reliable target parameters can be estimated for > 90% of anomalies on each sensor's detection list.	9 of 1282 (0.7%) classified as "Unable to analyze"
Correct estimation of target parameters	Accuracy of estimated target parameters.	• Demonstrator target parameters • Results of intrusive investigation	βs : COV < 0.20 X, Y < 15 cm (1σ) Z < 10 cm (1σ) Size: COV < 0.20	β1,2,3 COV=0.25,0.30,0.35 XY mean = 16cm (14cm σ) Z mean= -1cm (9cm σ) Size COV=0.10

Table 3-5 Performance Objectives for MetalMapper – 3 criteria

Performance Objective	Metric	Data Required	Success Criteria	Results
Quantitative Performance Objectives				
Maximize correct classification of munitions	Number of targets-of-interest retained.	• Prioritized anomaly lists • Scoring reports from IDA	Approach correctly classifies all targets-of-interest	200 of 204 (98%) correctly classified as TOI
Maximize correct classification of non-munitions	Number of false alarms eliminated.	• Prioritized anomaly lists • Scoring reports from IDA	Reduction of false alarms by > 30% while retaining all targets of interest	477 of 859 (56%) correctly classified as non-munitions
Specification of no-dig threshold	P_{class} and N_{fa} at demonstrator operating point.	• Demonstrator - specified threshold • Scoring reports from IDA	Threshold specified by the demonstrator to achieve criteria above	P_{class} = 0.98 N_{fa} = 220
Minimize number of anomalies that cannot be analyzed	Number of anomalies that must be classified as "Unable to Analyze."	• Demonstrator target parameters	Reliable target parameters can be estimated for > 90% of anomalies on each sensor's detection list.	1 of 1063 (0.1%) classified as "Unable to analyze"
Correct estimation of TOI target parameters	Accuracy of estimated target parameters.	• Demonstrator target parameters • Results of intrusive investigation	βs@0.3ms :COV<0.20 X, Y < 15 cm (1σ) Z < 10 cm (1σ) Size@0.3ms:COV<0.20	β1,2,3 COV=0.24,0.25,0.28 XY mean = 12cm (7cm σ) Z mean= 3cm (10cm σ) Size COV=0.08

3.1 OBJECTIVE: MAXIMIZE CORRECT CLASSIFICATION OF MUNITIONS

This is one of the two primary measures of the effectiveness of this approach. By collecting high-quality data and analyzing those data with advanced parameter estimation and classification algorithms we expect to classify the targets with high efficiency.

3.1.1 Metric

The metric for this objective is the number of items on the master anomaly list that were correctly classified as munitions by each classification approach.

3.1.2 Data Requirements

A prioritized dig list for the targets on the master anomaly list was prepared for each of the data sets analyzed as part of this demonstration. Using complete ground truth of all the anomalies on the master list IDA personnel used their scoring algorithms to assess the results.

3.1.3 Success Criteria and Results

The objective was considered to met if all of the items-of-interest were correctly labeled as munitions on the prioritized anomaly list.

The EM61 array data met the objective by classifying all 206 TOIs as anomalies to dig. Both the slope corrected (SCORR) and non-slope corrected (non-SCORR) EM61 cart data came close to meeting the objective by classifying 211 of 212 TOIs as anomalies to dig. The single failure was very close to the "dig/no dig" threshold. The EM61 MSEMS data classified 198 of 205 TOIs as anomalies to dig. The larger number of failures was due to an aggressive choice of the "dig/no dig" threshold which placed several 60mm mortars in the no dig category. The TEMTADS 2-criteria data classified 203 of 206 TOIs as anomalies to dig. The three TEMTADS failures were due to multiple objects and sampling problems. The Metal Mapper 3 criteria data classified 200 of 204 TOIs as anomalies to dig. The four Metal Mapper failures were caused by multiple objects, weak signals and UXO variability.

3.2 OBJECTIVE: MAXIMIZE CORRECT CLASSIFICATION OF NON-MUNITIONS

This is the second of the two primary measures of the effectiveness of this approach. By collecting high-quality data and analyzing those data with advanced parameter estimation and classification algorithms we expect to classify the targets with high efficiency. This objective targets the reduction of false alarms.

3.2.1 Metric

The metric for this objective is the number of items-of-interest on the master dig list that can be correctly classified as non-munitions by each classification approach.

3.2.2 Data Requirements

See section 3.1.2.

3.2.3 Success Criteria and Results

The objective was considered met if more than 30% of the non-munitions items were correctly labeled as non-munitions while retaining all of the targets-of-interest on the dig list.

The EM61 array, SCORR EM61 cart, non-SCORR EM61 cart and Metal Mapper met the objective by respectively labeling 52%, 55%, 55% and 56% of the non-munitions items as "no dig" without including any of the TOIs. The EM61 MSEMS was only able to remove 4% of the non-munitions items before encountering a TOI but this number increased to 54% before encountering another TOI. The problematic target was Master ID 444. There were several other anomalies that overlapped with this anomaly and multiple ground truth objects that caused problems with the analysis. The TEMTADS only removed 7% of the non-munitions items before encountering a TOI but this number increased to 60% before encountering another TOI. As with the EM61 MSEMS, the challenging anomaly contained multiple objects.

3.3 OBJECTIVE: SPECIFICATION OF NO-DIG THRESHOLD

In a retrospective analysis as was performed in this demonstration, it is possible to tell the true classification capabilities of a classification procedure based solely on the prioritized dig list submitted by each demonstrator. In a real-world scenario, all targets may not be dug so the success of the approach will depend on the ability of an analyst to accurately specify their dig/no-dig threshold.

3.3.1 Metric

P_{class} and number of false alarms, N_{fa}, at the demonstrator-specified threshold are the metrics for this objective.

3.3.2 Data Requirements

See section 3.1.2.

3.3.3 Success Criteria

The objective was considered met if more than 30% of the non-munitions items were correctly labeled as non-munitions while retaining all of the targets-of-interest at the demonstrator-specified threshold.

The EM61 array data met the objective by classifying all 206 TOIs as anomalies to dig while correctly labeling 49% of the non-munitions at the chosen threshold. Both the SCORR and non-SCORR EM61 cart data came close to meeting the objective by classifying 211 of 212 TOIs as anomalies to dig while correctly labeling 57% as non-munitions. The one failure was very close to the "dig/no dig" threshold. The EM61 MSEMS data classified 198 of 205 TOIs as anomalies to dig while correctly labeling 67% as non-munitions. While the EM61 MSEMS placed are higher percentage of non-munitions in the "no dig" category it also had the highest number of false negatives which points to an incorrect choice of the "dig/no dig" threshold. The TEMTADS data classified 203 of 206 TOIs as anomalies to dig while correctly labeling 70% as non-munitions. The Metal Mapper data classified 200 of 204 TOIs as anomalies to dig while correctly labeling 74% as non-munitions.

3.4 OBJECTIVE: MINIMIZE NUMBER OF ANOMALIES THAT CANNOT BE ANALYZED

Anomalies for which reliable parameters cannot be estimated cannot be classified by the classifier. These anomalies must be placed in the dig category and reduce the effectiveness of the classification process.

3.4.1 Metric

The number of anomalies for which reliable parameters cannot be estimated is the metric.

3.4.2 Data Requirements

A list of all target parameters along with a list of those anomalies for which parameters could not be reliable estimated was submitted for each of the data sets analyzed as part of this demonstration.

3.4.3 Success Criteria and Results

The objective was considered met if reliable parameters were estimated for > 90% of the anomalies on each sensor anomaly list.

All the sensors easily met this objective with target parameters calculated for 99-100% of the anomalies on each sensor anomaly list.

3.5 OBJECTIVE: CORRECT ESTIMATION OF TARGET PARAMETERS

This objective involved the accuracy of the target parameters that are estimated in the first phase of the analysis. Successful classification is only possible if the input features are internally consistent.

3.5.1 Metric

Accuracy of estimation of target parameters is the metric for this objective. The metric used for the XY location and depth accuracy is the mean and standard deviation of difference between the estimated and true values. The metric used for the estimation of size and betas is the coefficient of variation (COV), which is simply the standard deviation divided by the mean value of the parameter. It essentially reports variability in the estimated parameter. We used this normalized metric to compare performance because different EMI sensors return polarization estimates that vary in absolute magnitude.

3.5.2 Data Requirements

A list of all target parameters were submitted for each of the data sets analyzed as part of this demonstration. SAIC analysts compared these estimated parameters for all TOI with good isolated or marginally overlapping signals to those measured during the intrusive investigation.

3.5.3 Success Criteria and Results

The objective was considered met if the COV for the estimated polarizations are less than .20, the estimated X, Y locations are within 15 cm (1σ), and the estimated depths are within 10 cm (1σ).

In general, the EM61 sensors were able to meet the target parameter objectives for size and XY location but failed to meet them for Z and individual betas with a few exceptions. As expected the slope of the terrain in the survey area had an effect on the final XY position. The EM61 cart without the slope correction failed the objective and had the worst accuracy with a mean error of 31cm and a standard deviation of 17cm. In comparison, the slope corrected EM61 cart data had a mean error of 22cm with a standard deviation of 13cm.

The cued array systems produced more accurate target parameters than the dynamic systems. Both the TEMTADS and the Metal Mapper met the objectives for size, XY and Z but did not quite meet the objectives for the individual betas. This was due to the variability of the TOI at the site. The statistics were calculated individually for the four main TOI found and averaged to generate the final COV value. On its own the 4.2inch mortar would pass the objective because of the high amplitude signal and minimal target variability. On the other hand the low amplitude signal and target variability (body only, complete, seeded and native) of the 60mm mortars produced a COV value above the objective threshold. A lower COV value would be calculated if each of the 60mm mortars were further divided into subclasses.

4 SITE DESCRIPTION

The site description material reproduced here is taken from the recent SI report [8]. The former Camp San Luis Obispo is approximately 2,101 acres situated along Highway 1, approximately five miles northwest of San Luis Obispo, California. The majority of the area consists of mountains and canyons. The site for this demonstration is a mortar target on hilltop in Munitions Response Site (MRS) 05 (within former Rifle Range #12).

4.1 SITE SELECTION

This site was chosen as the next in a progression of increasingly more complex sites for demonstration of the classification process. The first site in the series, Camp Sibert, had only one target-of-interest and item "size" was an effective discriminate. At this site, there were four main targets-of-interest: 60-mm, 81-mm, 4.2-in mortars and 2.36-in rockets. During the excavation phase of the demonstration the following additional TOIs were found: 37mm projectile, 3inch stokes mortar and 5inch rocket warhead.

4.2 SITE HISTORY

Camp San Luis Obispo was established in 1928 by California as a National Guard Camp. Identified at that time as Camp Merriam, it originally consisted of 5,800 acres. Additional lands were added in the early 1940s until the acreage totaled 14,959. During World War II, Camp San Luis Obispo was used by the U.S. Army from 1943 to 1946 for infantry division training including artillery, small arms ranges, mortar, rocket, and grenade ranges. According to the Preliminary Historical Records Review (HRR), there were 27 ranges and thirteen training areas located on Camp San Luis Obispo during World War II. Construction at the camp included typical dwellings, garages, latrines, target houses, repair shops, and miscellaneous range structures. Following the end of World War II, a small portion of the former camp land was returned to its former private owners. The U.S. Army was making arrangements to relinquish the rest of Camp San Luis Obispo to the State of California and other government agencies when the conflict in Korea started in 1950. The camp was reactivated at that time.

The U.S. Army used the former camp during the Korean War from 1951 through 1953 where the Southwest Signal Center was established for the purpose of signal corps training. The HRR identified eighteen ranges and sixteen training areas present at Camp San Luis Obispo during the Korean War. A limited number of these ranges and training areas were used previously during World War II. Following the Korean War, the camp was maintained in inactive status until it was relinquished by the Army in the 1960s and 1970s. Approximately 4,685 acres was relinquished to the General Services Administration (GSA) in 1965. GSA then transferred the property to other agencies and individuals beginning in the late-1960s through the 1980s; most of which was transferred for educational purposes (Cal Poly and Cuesta College). A large portion of Camp San Luis Obispo (the original 5,880 acres) has been retained by the California National Guard (CNG) and is not part of the FUDS program.

4.3 SITE TOPOGRAPHY AND GEOLOGY

The Camp San Luis Obispo site consisted mainly of mountains and canyons classified as grassland, wooded grassland, woodland, or brush. A major portion of the site is identified as grassland and is used primarily for grazing. Los Padres National Forest (woodland) is located to the north-northeastern portion of the site.

The underlying bedrock within the Camp San Luis Obispo site area is intensely folded, fractured, and faulted. The site is underlain by a mixture of metamorphic, igneous, and sedimentary rocks less than 200 million years old. Scattered throughout the site are areas of fluvial sediments overlaying metamorphosed material known as Franciscan mélange. These areas are intruded by plugs of volcanic material that comprise a chain of former volcanoes extending from the southwest portion of the site to the coast. Due to its proximity to the tectonic interaction of the North American and Pacific crustal plates, the area is seismically active.

A large portion of the site consists of hills and mountains with three categories of soils occurring within: alluvial plains and fans; terrace soils; and hill/mountain soils. Occurring mainly adjacent to stream channels are the soils associated with the alluvial plains and fans. Slope is nearly level to moderately sloping and the elevation ranges from 600 to 1,500 feet. The soils are very deep and poorly drained to somewhat excessively drained. Surface layers range from silty clay to loamy sand. The terrace soils are nearly level to very steep and the elevations ranges from 600 to 1,600 feet. Soils in this unit are considered shallow to very deep, well drained, and moderately well drained. The surface layer is coarse sandy loam to shaley loam. The hill/mountain soils are strongly sloping to very steep. The elevation ranges from 600 to 3,400 feet. The soils are shallow to deep and excessively drained to well drained with a surface layer of loamy sand to silty clay.

4.4 MUNITIONS CONTAMINATION

A large variety of munitions have been reported as used at the former Camp San Luis Obispo. Munitions debris from the following sources was observed in MRS 05 during the 2007 SI:

- 4.2-inch white phosphorus mortar
- 4.2-inch base plate
- 3.5-inch rocket
- 37mm
- 75mm
- 105mm
- 60mm mortar
- 81mm mortar
- practice bomb
- 30 cal casings and fuzes.

- flares found of newer metal; suspected from CNG activities

At the particular site of this demonstration, 60-mm, 81-mm, and 4.2-in mortars and mortar fragments are known to be present.

4.5 SITE GEODETIC CONTROL INFORMATION

Table 4-1. Geodetic Control at the former Camp San Luis Obispo site

ID	Latitude	Longitude	Elevation (m)	Northing (m)	Easting (m)	HAE (m)
ESTCP	35º 20' 37.77465" N	120º 44' 25.95073"W	113.69	3,913,515.94	705,330.89	76.01

4.6 SITE CONFIGURATION

The demonstration site was configured as one 11.8-acre area. It spanned the hillside that is the historical mortar target. The cart systems surveyed 45 30- x 30-m grids within this area for a total of 10 acres. The vehicular systems surveyed the entire area. Details of the final site extent are shown in Figure 4-1. The calibration strip and training pit were located off the site, convenient to the access road.

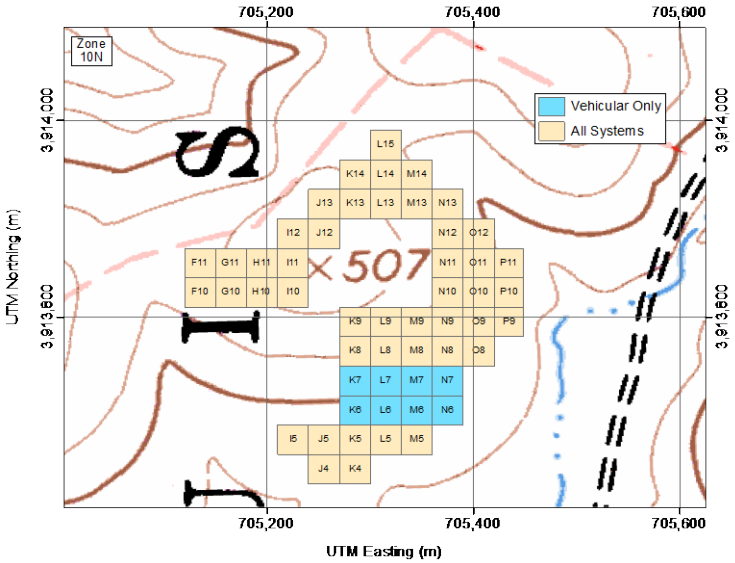


Figure 4-1. Final layout of the demonstration site showing the grids surveyed by all systems (10 acres) and the additional 8 grids (1.8 acres) surveyed by the vehicular systems.

5 TEST DESIGN

5.1 CONCEPTUAL EXPERIMENTAL DESIGN

The overall objective of the ESTCP program was to demonstrate a methodology for the use of classification in the munitions response process. The three key components of this methodology were collection of high-quality geophysical data and principled selection of anomalous regions in those data, analysis of the selected anomalies using physics-based models to extract target parameters such as size, shape, and materials properties, and the use of those parameters to construct a prioritized dig list. Each of these components was handled separately in this program with SAIC focusing on the last two objectives.

The ESTCP Program Office coordinated data collection activities. This included all preparatory activities, arranging for a data collection by well-validated systems, selection of anomalies for analysis from each geophysical data set, and compilation of the individual sensor anomaly lists into a master list.

SAIC then processed the individual data sets using existing routines to extract target parameters. These parameters were passed to the classification routines which, after training on a limited amount of site-specific ground truth, were used to produce prioritized dig lists.

Validation digging was coordinated by the Program Office. Since this is a demonstration, all anomalies on the master dig list were investigated. The underlying target was uncovered, photographed, located with a cm-level GPS system, and removed. The identities of a small number of the recovered items were provided to SAIC for use as training data. These data were used as inputs to finalize algorithms and adjust thresholds.

At the conclusion of training, SAIC submitted eight prioritized dig lists. Four dig lists were based on UX-Analyze processing of the survey data sets; including, MTADS EM61 array, Contractor EM61 cart (with and without slope correction) and MSEMS EM61. The remaining four dig lists were submitted for the TEMTADS and Metal Mapper data. These lists were ordered from the item that is most confident not hazardous through the item that is most confident munitions. The anomalies for which we were not able to extract meaningful parameters will be placed at the bottom of the list. These inputs were scored by the Institute for Defense Analyses (IDA) with emphasis on the number of items that were correctly labeled non-hazardous while correctly labeling all munitions items.

The primary objective of the demonstration was to assess how well we are able to order the prioritized anomaly list and specify the threshold separating high confidence clutter from all other items.

5.2 PRE-DEMONSTRATION ACTIVITIES

Pre-demonstration activities performed by ESTCP can be found in Section 5 of the ESTCP UXO Discrimination Study Demonstration Plan [3].

5.3 CALIBRATION ACTIVITIES

Details of the sensor calibration measurements can be found in Section 5 of the ESTCP UXO Discrimination Study Demonstration Plan [3].

5.4 DATA COLLECTION PROCEDURES

5.4.1 Scale of Demonstration

The demonstration was conducted at the Former Camp San Luis Obispo, CA.

5.4.2 Sample Density

Survey data with the commercial EMI instruments were collected with a line spacing of 0.5-m. The EM61-MK2 array collected data in two orthogonal directions doubling the sample density. The magnetometer array data was collected with 0.25-m line spacing while the MSEMS magnetometer data was collected on 0.5-m spacing.

5.4.3 Quality Checks

Details of data quality checks and the twice-daily surveys of the calibration strip can be found in the individual demonstrator demonstration plans [4-7].

5.4.4 Data Handling

Each data collection demonstrator provided raw and located, preprocessed data to the ESTCP Program Office for archiving. For those systems without built-in sensor orientation measurement (EM61-MK2 cart and MSEMS) the Program Office team corrected the initial sensor locations using a digital slope model developed from the EM array data acquired by the Naval Research Laboratory. These "slope-corrected" data were used for all subsequent analysis. In addition to the slope corrected (SCORR) data the original data from the EM61-MK2 cart was analyzed for comparison purposes.

5.5 ANOMALY SELECTION

Anomalies were selected from each geophysical data set by the data collectors using a target response-based threshold. An example of the response of an EM61-MK2 cart to a 60-mm mortar as a function of depth is shown in Figure 5-1.

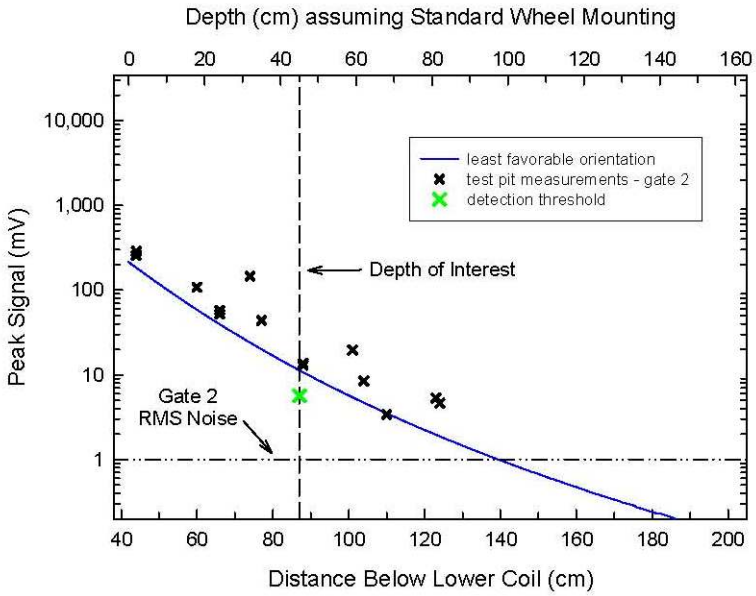


Figure 5-1. Calculated EM61-MK2 cart signal expected from a 60-mm mortar when the mortar is in its least favorable orientation. Actual survey measurements from a training pit are shown as validation for the predictions. The depth corresponding to 11x the mortar's diameter is marked.

In conjunction with the Advisory Group for the Demonstration, the ESTCP Program Office set the depth of interest for all items at this site as 45 cm below ground (depth measured to the center of the object). The anomaly selection threshold was established by adjusting the minimum signal expected at this depth downward by a safety factor (50% for this program) as shown in the figure. Since there are four targets-of-interest are expected at this site, the ultimate anomaly selection threshold for each data set was the smallest of the four individual item thresholds.

5.6 CONSTRUCTION OF A MASTER ANOMALY LIST

IDA combined the different sensor-specific target lists into a Master Anomaly List [9].

5.7 CUED DATA COLLECTION

The NRL TEMTADS system worked from a cued list that was constructed by IDA. In contrast, the Metal Mapper cued-mode anomaly list was derived from the dynamic Metal Mapper survey.

5.8 VALIDATION

At the conclusion of data collection activities, all anomalies on the master anomaly list were excavated. Each encountered item was identified, photographed, its depth measured, its location determined using cm-level GPS, and removed if possible. All non-hazardous items were saved for subsequent in-air measurements.

6 DATA ANALYSIS PLAN

6.1 PREPROCESSING

Survey data were preprocessed (located and simple filtering) by the data collection demonstrators in preparation for anomaly selection [4-7].

6.2 PARAMETER ESTIMATION

6.2.1 Magnetic Data

The MTADS magnetometer array data and the MSEMS magnetometer data were analyzed using UX-Analyze, whose inversion routines assume a dipole source. The magnetic dipole model inverts for the location (X_0, Y_0, Z_0), magnetic moment, and orientation angles (*declination and inclination*). The model first fits the shape of the footprint, then its magnitude. Initial guesses for the fit parameters are determined internally within the code based on the measured signature. The goodness of fit metric, ε, is the squared correlation of data and model. By definition, the dipole fit error metric is equal to $\sqrt{1-\varepsilon}$.

6.2.2 EM61 Data

The MTADS EM61 array, contractor EM61 cart and MSEMS EM61 data were also analyzed using the characterization algorithms found in UX-Analyze. The EMI data reflect details of the sensor/target geometry as well as inherent EMI response characteristics of the targets themselves. In order to separate out the intrinsic target response properties from sensor/target geometry effects we invert the signature data to estimate principal axis magnetic polarizabilities for the targets. The EMI data are inverted using the standard induced dipole response model wherein the effect of eddy currents set up in the target by the primary field is represented by a set of three orthogonal magnetic dipoles at the target location [10,11]. The measured signal is a linear function of the induced dipole moment **m**, which can be expressed in terms of a time dependent polarizability tensor **B** as

$$\mathbf{m} = \mathrm{UBU}^{\mathrm{T}}\mathbf{H}_0$$

where $\mathbf{H}_0$ is the peak primary field at the target, U is the transformation matrix between the coordinate directions and the principal axes of the target, and B is an empirically determined, effective magnetic polarizability matrix. For an arbitrary compact object, this matrix can be diagonalized about three primary body axes and written as:

$$\mathrm{B} = \begin{bmatrix} \beta_X & 0 & 0 \\ 0 & \beta_Y & 0 \\ 0 & 0 & \beta_Z \end{bmatrix}.$$

Given a set of measurements of the target response with varying geometries or "look angles" at the target, the data can be inverted to determine the (X, Y, Z) location of the target, the

orientation of its principal axes (ψ, θ, ϕ), and the principal axis polarizabilities (β_x, β_y, β_z). The basic idea is to search out the set of nine parameters (X, Y, Z, ψ, θ, ϕ, β_x, β_y, β_z) that minimizes the difference between the measured responses and those calculated using the dipole response model. The model used in UX-Analyze applies a non-linear Marquardt-Levenberg search for all nine parameters.

The relative magnitudes of the β's are determined by the size, shape and composition of the object as well as the transmit waveform and time gate or frequency. The transformation matrix contains the angular information about the orientation of these body axes.

For cylindrical objects like most UXO, B is a diagonal matrix with only two unique coefficients, corresponding to the longitudinal (β_T) and transverse (β_L) directions:

$$B = \begin{bmatrix} \beta_L & 0 & 0 \\ 0 & \beta_T & 0 \\ 0 & 0 & \beta_T \end{bmatrix}.$$

Discrimination is based on target β's estimated from spatially mapped data. Specific ordnance items have specific β values, while clutter items generally have different β values.

6.2.3 TEMTADS and Metal Mapper Data

The TEMTADS array and Metal Mapper data uses the same dipole model as that described above for the EM61 data but the inversion is accomplished by a two-stage method. In the first stage, the target's (X, Y, Z) dipole location beneath is solved for non-linearly. At each iteration within this inversion, the nine element polarizability tensor (**B**) is solved linearly. We require that this tensor be symmetric; therefore, only six elements are unique. Initial guesses for X and Y were determined by a signal-weighted mean. The routine loops over a number of initial guesses in Z, keeping the result giving the best fit as measured by the chi-squared value. The non-linear inversion is done simultaneously over all time gates, such that the dipole location applies to all decay times. At each time gate, the eigenvalues and angles are extracted from the polarizability tensor.

In the second stage, six parameters are used: the three spatial parameters (X, Y, Z) and three angles representing the yaw, pitch, and roll of the target (Euler angles ψ, θ, ϕ). Here the eigenvalues of the polarizability tensor are solved for linearly within the 6-parameter non-linear inversion. In this second stage both the target location and its orientation are required to remain constant over all time gates. The value of the best fit X, Y, and Z from the first stage, and the median value of the first-stage angles are used as an initial guess for this stage. Additional loops over depth and angles are included to better ensure finding the global minimum.

Figure 6-1 shows an example of the principal axis polarizabilities determined from TEMTADS array data. The target, a mortar fragment, is a slightly bent plate about ½ cm thick, 25 cm long and 15 cm wide. The red curve is the polarizability when the primary field is normal to the

surface of the plate, while the green and blue curves correspond to cases where the primary field is aligned along each of the edges.

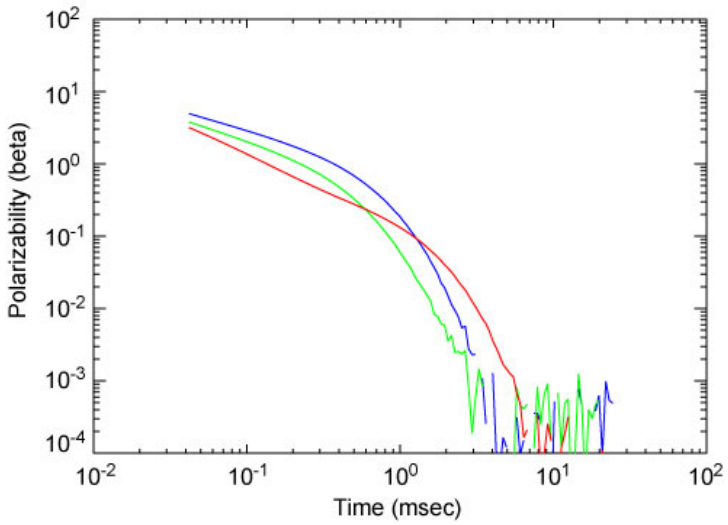


Figure 6-1. Principal axis polarizabilities (units of m^3) for a ½ cm thick by 25cm long by 15cm wide mortar fragment.

6.3 TRAINING AND CLASSIFICATION

All of the classification approaches required some level of training data. These data came from (i) sensor data for the targets-of-interest collected in previous testing, (ii) data collected over the training pit and calibration line, and (iii) ground truth from excavations of the first five grids.

Specific training and classification procedures for each data set or processing approach are discussed individually in the sections that follow.

SAIC submitted prioritized dig lists using the structure and nomenclature shown Figure 6-2.

Anomaly ID	Decision Statistic	Rank	Category	Overlaps	Comments
247	.97	1	1		
1114	.96	2	1		High confidence NOT munition
69	…	3	1		
…	…	…			
			Threshold		
…	…	…	2	1	
…	…	…	2		Can't make a decision
…	…	…	2	1	
…	…	…	2		
…	…	…	3	2	
…	…	N-2	3	2	High confidence munitions
…	.03	N-1	3		
…	.02	N	3		
…		…			
…	-9999	-9999	4		
…	-9999	-9999	4		Can't extract reliable features
…	-9999	-9999	4		

Figure 6-2. Format for the prioritized anomaly list that was submitted for each classification approach

6.3.1 Magnetic Data

The MTADS magnetometer array data and MSEMS magnetometer data were analyzed using UX-Analyze. There were 182 training anomalies for both magnetic data sets but because the anomaly selections were based on the EM61 Array data many of the selections were either very weak or did not contain an anomalous response. There were also several magnetic anomalies in the survey area, which were much too large for the expected ordnance items at this site that hindered the analysis of several of the targets. For these reasons, it was decided by the Program Office not to try to discriminate targets of interest using only the magnetic data. Instead the magnetometer data were analyzed to see if they could be combined with the EMI data to discriminate TOI from clutter better than using only the EMI data.

6.3.1.1 MTADS Magnetometer Data

Dipole inversion results were obtained for only 141 of the 182 training targets. The majority of the targets that did not fit were because the anomalies were either too weak or did not exist at the target location. Of the 141 targets that produced fit parameters only 120 matched targets from the EM61 cart training list. These anomalies were visually inspected and only isolated targets or slightly overlapping targets with dipole fit errors less than 0.45 were selected for discrimination analysis. This resulted in 93 good training targets. Figure 6-3 plots the fitted size versus the angle between the targets magnetic dipole and the Earth's field (solid angle) for all the selected targets. Each of the TOI and the non-TOI are plotted as different color symbols. As evident by

the figure the fitted size of the TOI overlap and cover most of the feature space. Only targets with a size less than about 0.025 fall outside the range covered by the TOI. Likewise the solid angle of the TOI also cover most of the feature space. There was a 2.36" rocket with a solid angle of about 100° in the calibration strip that forced a higher than desired threshold.

Using an aggressive size threshold of 0.025 resulted in about 10 targets that can be classified as high confidence clutter using the magnetic data. Of these 10 targets only 2 were not classified as high confidence clutter using only the EM61 cart data. For this reason we concluded that for this site the effort involved to analyze the magnetic data is not worth the very small gain in the high confidence clutter category. With approval from the Program Office it was decided to not produce a cooperative dig list using the MTADS magnetometer data.

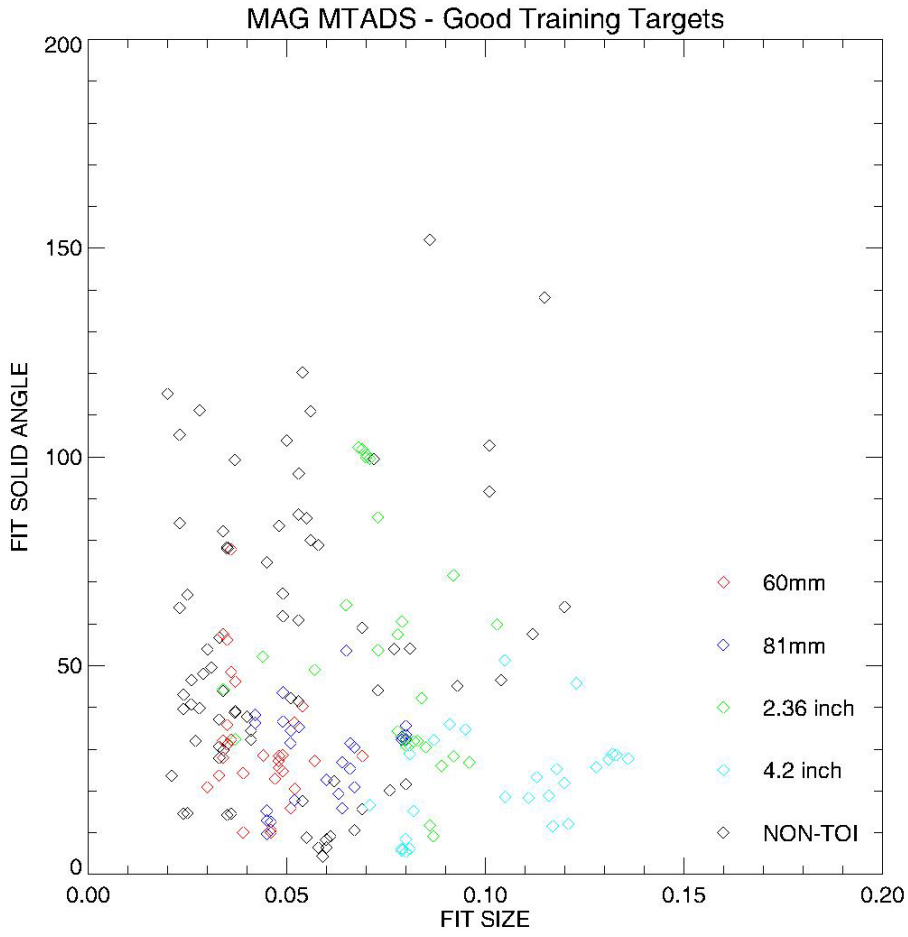


Figure 6-3. Plot of the fit size versus the inverted solid angle between the dipole moment and the Earth's field for the MTADS magnetometer array data.

6.3.1.2 MSEMS Magnetometer Data

The MSEMS magnetometer produced fewer valid targets compared to the MTADS magnetometer data. Only 87 of the initial 182 training targets produce dipole inversion results. The majority of the targets that did not fit were also because the anomalies were either too weak or did not exist at the target location. The increased number of targets that did not fit can be

attributed to the difference in sensor height (0.5m for MSEMS versus 0.33m for MTADS) and the sensor spacing (0.5m for MSEMS versus 0.25 for MTADS). Comparing these targets to the training targets selected for the EM61 MSEMS data resulted in only 78 common targets. Of these 78 targets only 62 passed our criteria of using only isolated targets or slightly overlapping targets with dipole fit errors less than 0.45. Figure 6-4 plots the fitted size versus the solid angle for all the selected MSEMS magnetometer targets. Each of the TOI and the non-TOI are plotted as different color symbols. Similar to the MTADS data, the figure shows the fitted size of the TOI overlapping and covering most of the feature space. Only targets with a size less than about 0.025 fall outside the range covered by the TOI. As with the MTADS magnetometer data only a few targets contained solid angles greater than the cutoff required to include all the TOI in the training data.

Using a size threshold of 0.025 resulted in only a few targets being classified high confidence clutter. All of these targets, however, were also classified as high confidence clutter using only the EM61 MSEMS data. Including the MSEMS magnetometer data with the EM61 MSEMS does not improve the discrimination performance compared to using only the EM61 MSEMS data. With approval from the Program Office it was decided to not produce a cooperative dig list using the MSEMS magnetometer data.

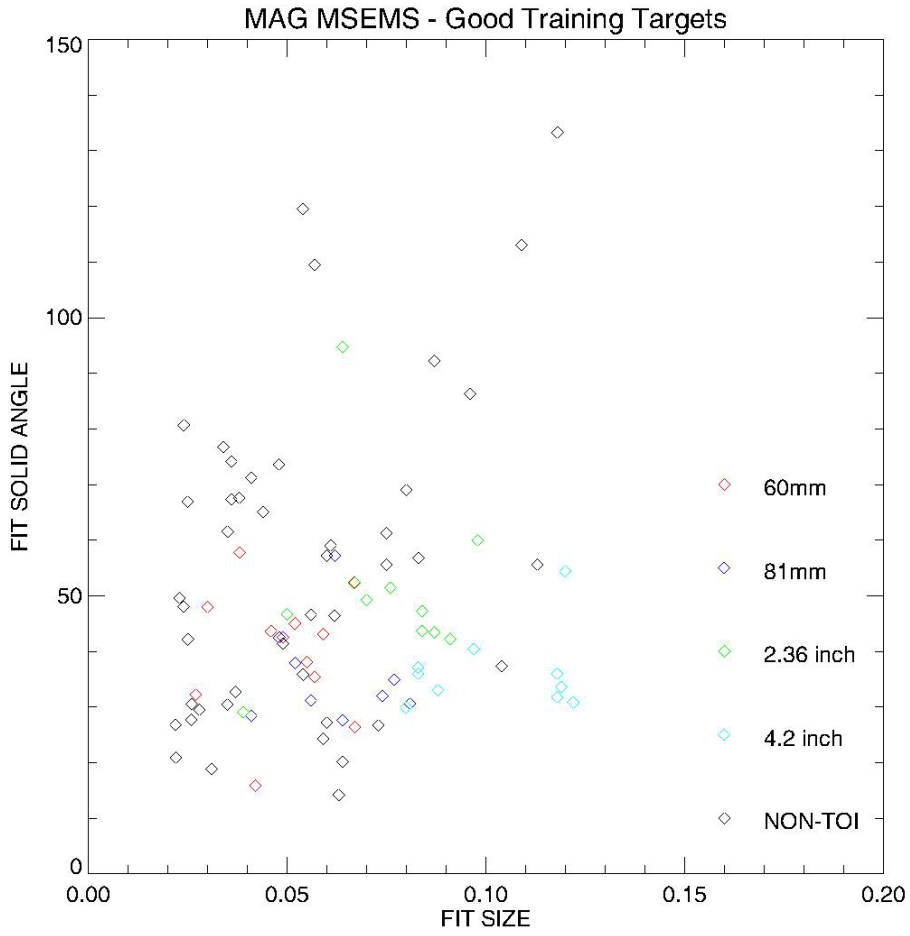


Figure 6-4. Plot of the fit size versus the inverted solid angle between the dipole moment and the Earth's field

6.3.2 EM61 Data

Multiple sets of EM61 data were collected at Camp SLO. These include EM61 array data, EM61 cart data, and EM61 data collected as part of the MSEMS platform. In partial fulfillment to MM-0704, we used UX-Analyze to characterize and classify the three EM61 data sets. For the EM61 cart data, we analyzed the data before and after slope corrections were made. As a result, we submitted four EM61 prioritized dig lists.

Scatter plots are presented below to provide an overall feel for the inverted target attributes. More complete attribute information for each data set is shown later in this section.

Figure 6-5 shows inverted polarizabilities EM61 cart data. Colors are used to identify labeled data. As shown here, the inverted shape attributes possess some discrimination information, but the intraclass variance and interclass separation is not ideal. The EM61 array data and MSEMS data showed similar results.

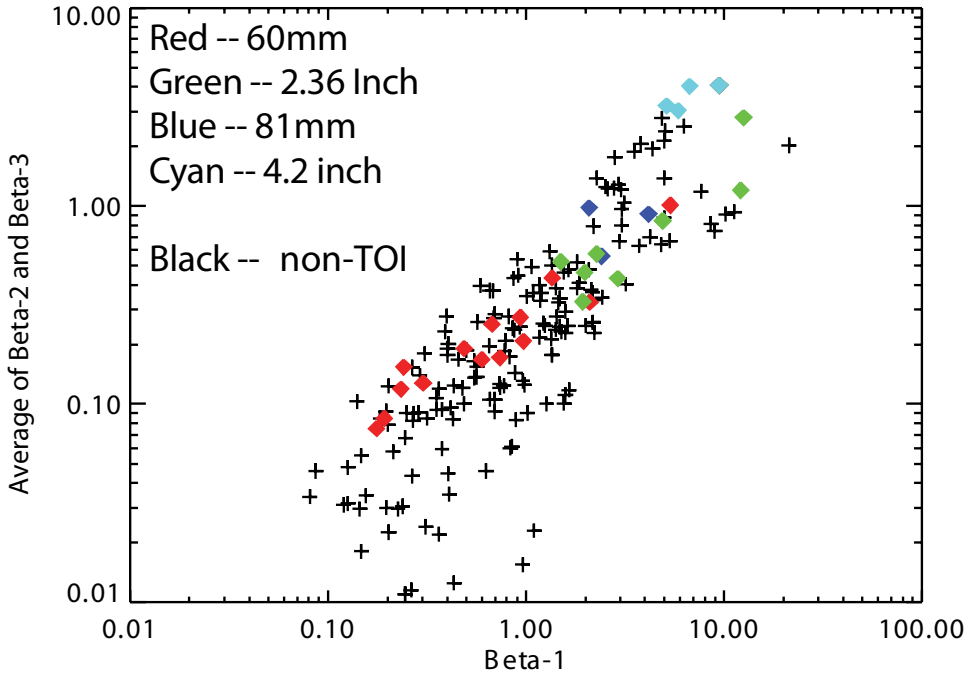


Figure 6-5. Principle axis polarizations for the labeled TOI, EM61 cart data (uncorrected for slope).

Figure 6-6 presents the inverted fit size versus the decay ratio for anomalies with a QC rank of good (QC process described later). The size metric can be used to distinguish 60mm from 4.2 mortars, but the 81mm and 2.36mm mortars cannot be easily separated using fitted size metric.

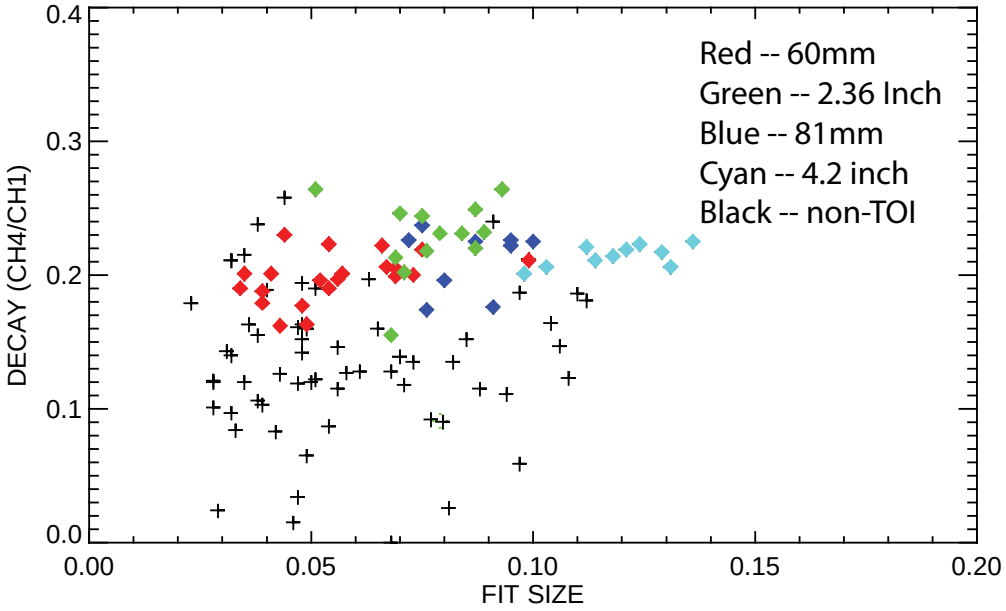


Figure 6-6. Inverted size parameter versus observed time decay for labeled TOI, EM61 cart data (slope corrected).

The decay information, however, does a decent job of discriminating between most of the TOI and most of the non-TOI. Figure 6-7 presents EM61 cart data (slope corrected, all data). Here we plot the fitted size versus the decay as measured by a fourth to first time gate (1266 and 216 µsec, respectively) average for measurements near that targets' peak signal.

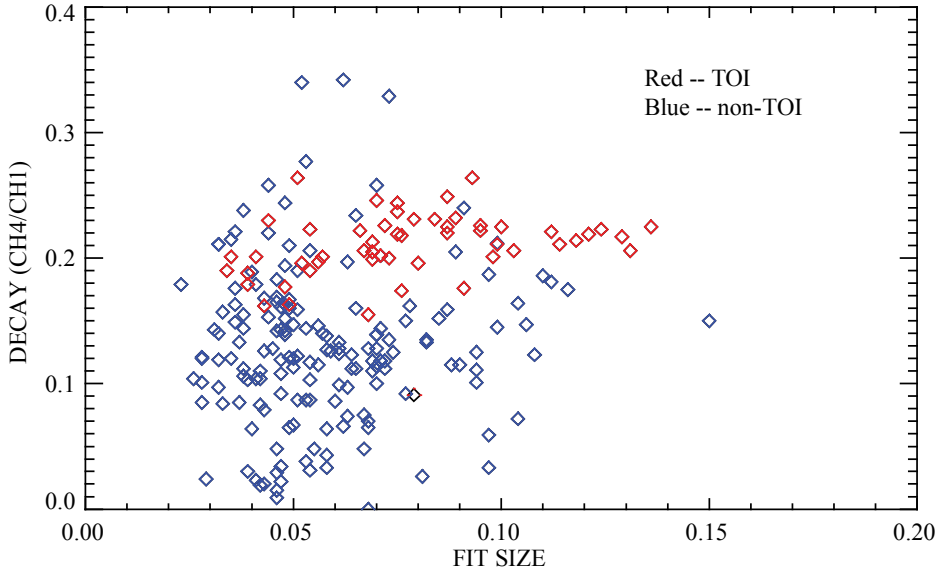


Figure 6-7. Scatter plot showing the fitted size versus decay ratio for EM61 cart data.

We utilized a Generalized Likelihood Ratio Test (GLRT) classifier on multiple target attributes and found the best performance was obtained when we used the inverted size and extracted decay rates.

The final classification was a straight-forward, multi-step process. For each anomaly, we calculated the GLRT probability that the test item was each of the TOI types individually and kept the target type that produced the highest confidence value. If the data were of good quality and the fit error was low, we simply took this maximum probability as the decision statistic. If the fit error was larger than 40%, we scaled and combined the individual probabilities. If the anomaly was poorly sampled and did not contain nine measurements over its peak, we assigned a probability of 0.5.

Various data scatter plots and performance metrics for each data set under investigation are shown in Figure 6-8 to Figure 6-23 and Table 6-1 to Table 6-4.

The first series of plots for each data type (indicated on the page in the upper right hand; i.e., "1/4" indicates the first of four pages for that particular data) shows the scatter in the positioning, fitted versus ground truth depth, and fit error as a function of signal strength.

The second page of plots for each data type plots the fitted size versus decay. The top segments by TOI vs non-TOI, the middle highlights fitted size of various TOI types, and the bottom plot indicates which of the labeled data were actually used for training (some were not used due to data quality issues).

The third plot page for each data type shows the decision metric. Here we plot the decision metric for each labeled anomaly. Colors are used to indicate rank as indicated in the figure caption. Because these data types were acquired using EM61 sensors, the thresholds are nominally consistent. They deviate slightly because of data quality issues of individual anomalies and the fact that all three sensors did not evaluate a common data set. In general, however, the decision metric threshold between Category 1 and 2 (the dig / do not dig threshold) was 0.2 (±0.1). The threshold between Category 2 and Category 3 was 0.37 (±0.03). A table showing the breakdown of the training data obtained from excavation of the first five grids is also displayed.

The fourth plot-page for each data type shows performance using a ROC curve. The top and bottom figures are identical except for color. The top display is in black and white, and the bottom graph is colored according to category.

EM61 CART – SLOPE CORRECTION APPLIED (1/4)

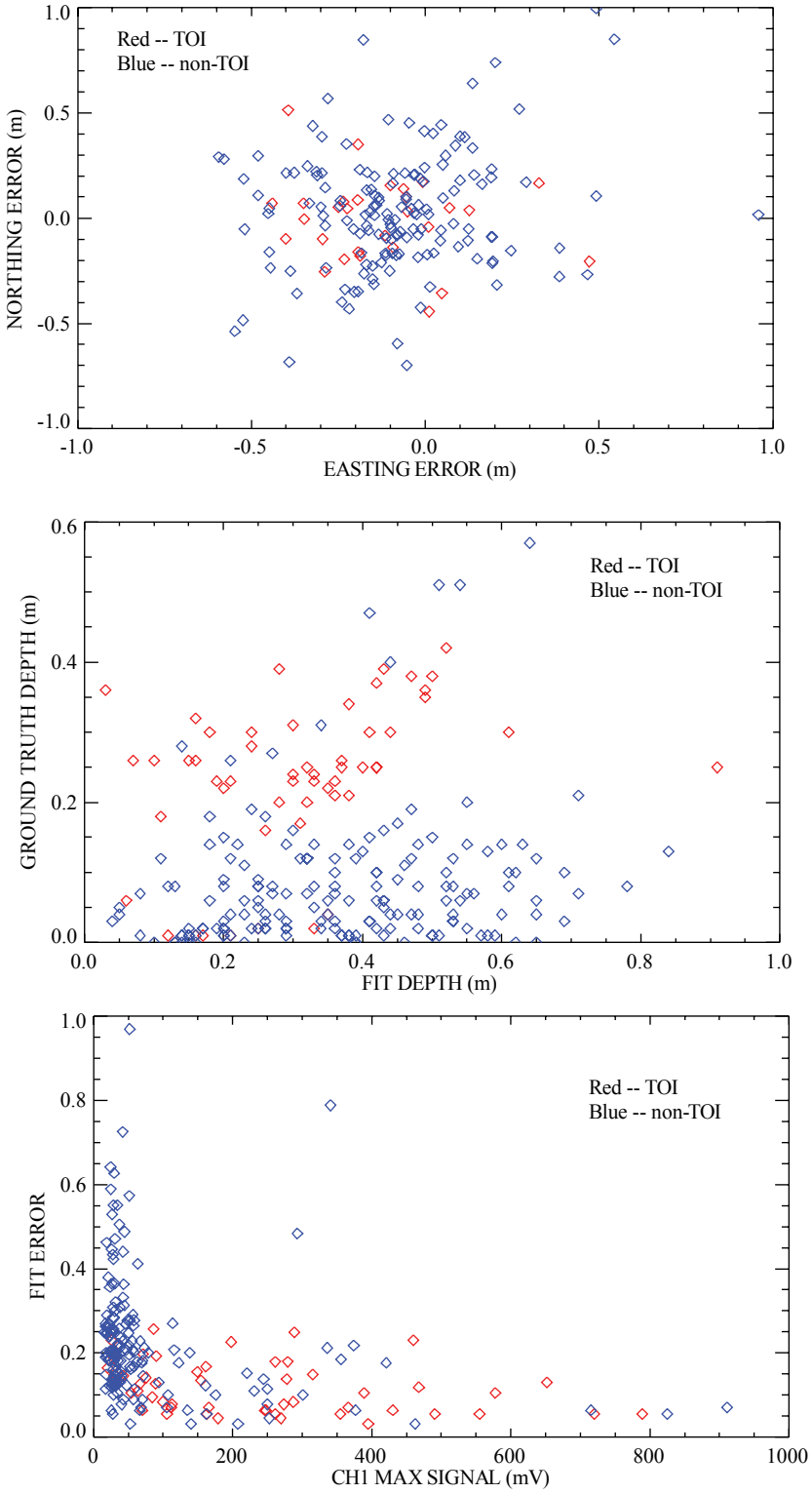


Figure 6-8. Plots showing scatter in the positioning, fitted versus ground truth depth, and fit error as a function of signal strength for the EM61 cart with slope correction data.

EM61 CART – SLOPE CORRECTION APPLIED (2/4)

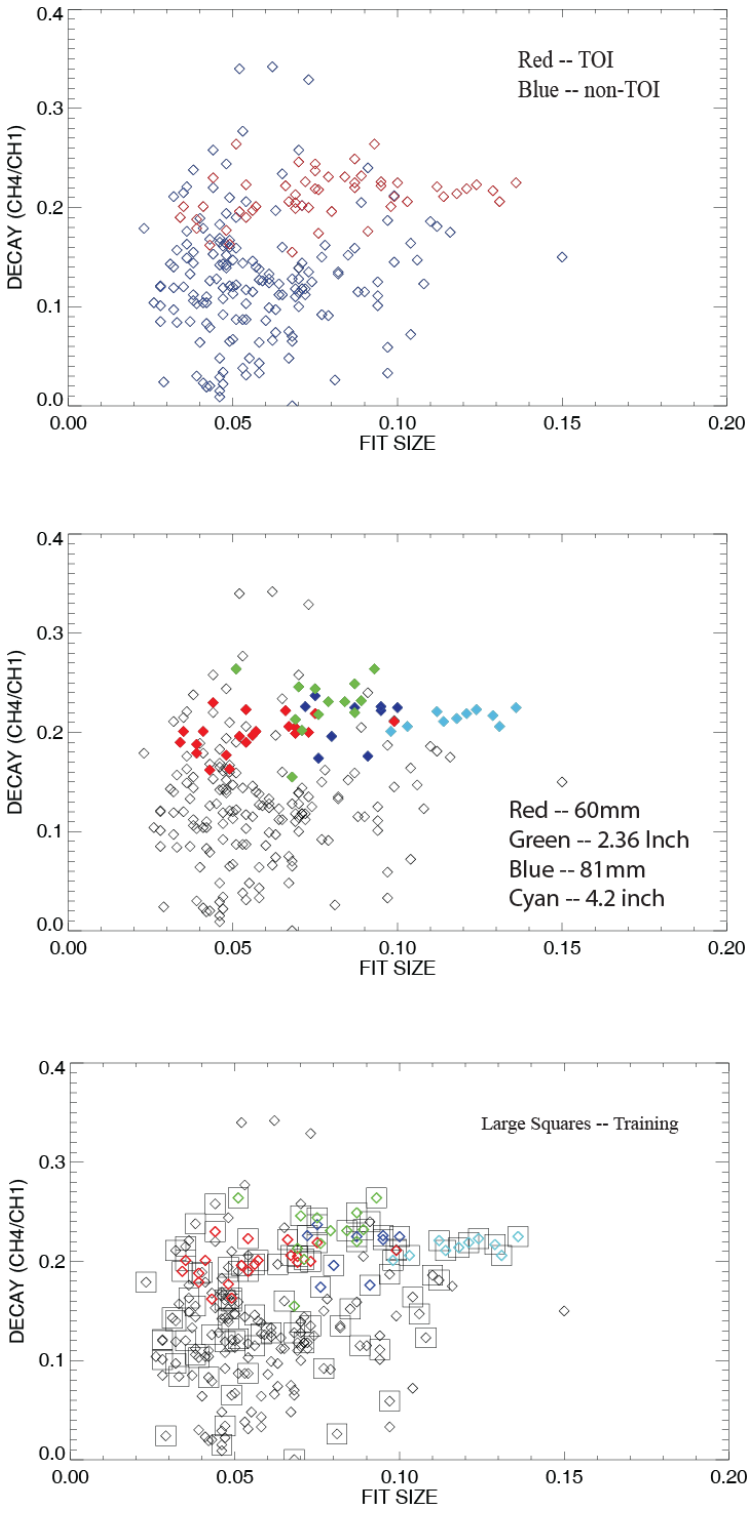


Figure 6-9. Plots of the fitted size versus decay. The three plots highlight different aspects of the training data for the EM61 cart with slope correction.

EM61 CART – SLOPE CORRECTION APPLIED (3/4)

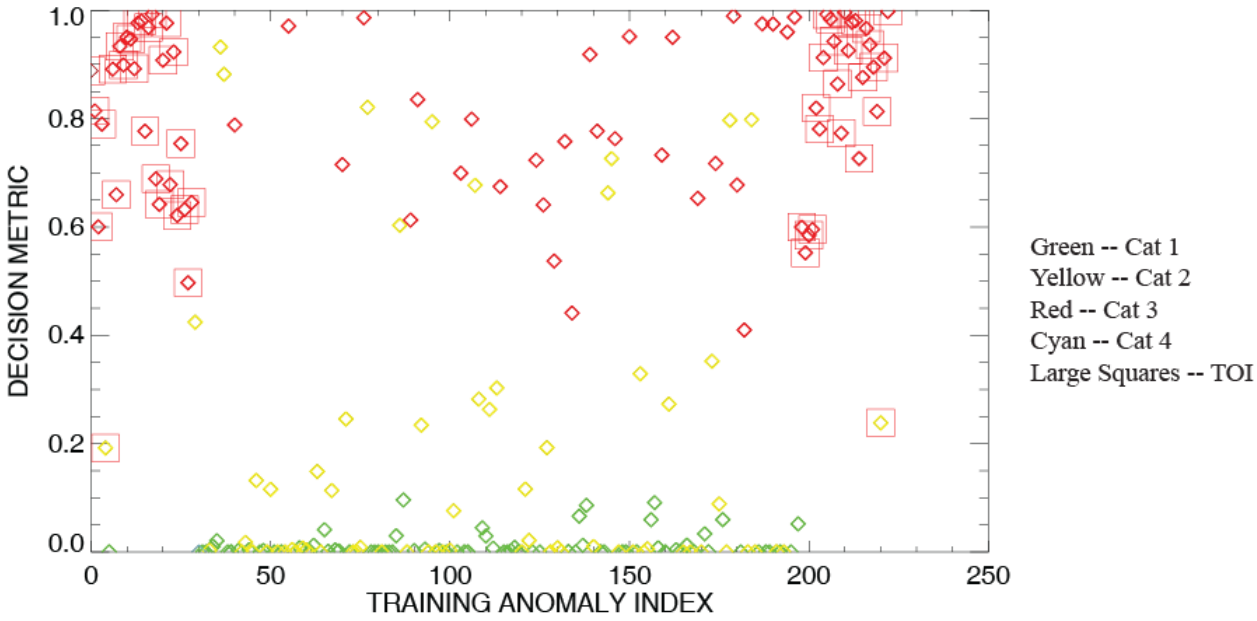


Figure 6-10. Plot of the decision metric for the EM61 cart with slope correction data. Colors are used to indicate rank as indicated in the figure caption.

Thresholds used for drawing ROC curve (next page):

Threshold separating Category 1 and 2: 0.1

Threshold separating Category 2 and 3: 0.4

Table 6-1 Training Set Summary: Slope corrected EM61 MK2 Cart

	Cultural	Munition Debris	No Contact	Soil	UXO
1	6	68	0	0	0
2	4	60	0	1	1
3	2	27	0	0	27
4	0	0	0	1	0
TOTAL	**12**	**155**	**0**	**2**	**28**

EM61 CART – SLOPE CORRECTION APPLIED (4/4)

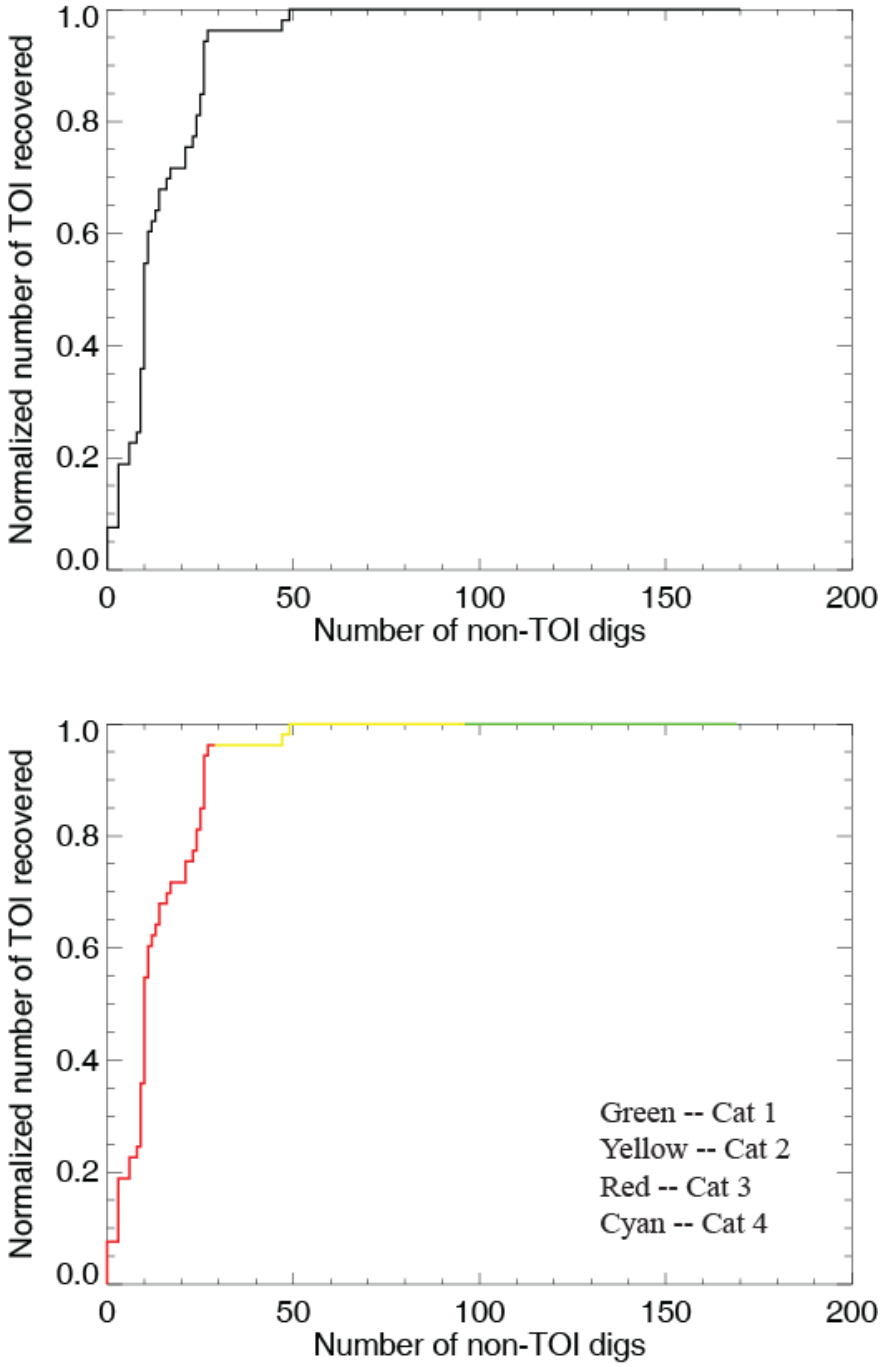


Figure 6-11. ROC curve of the EM61 cart with slope correction training data. The top and bottom figures are identical except that the bottom graph is colored according to category.

EM61 CART – NO SLOPE CORRECTION APPLIED (1/4)

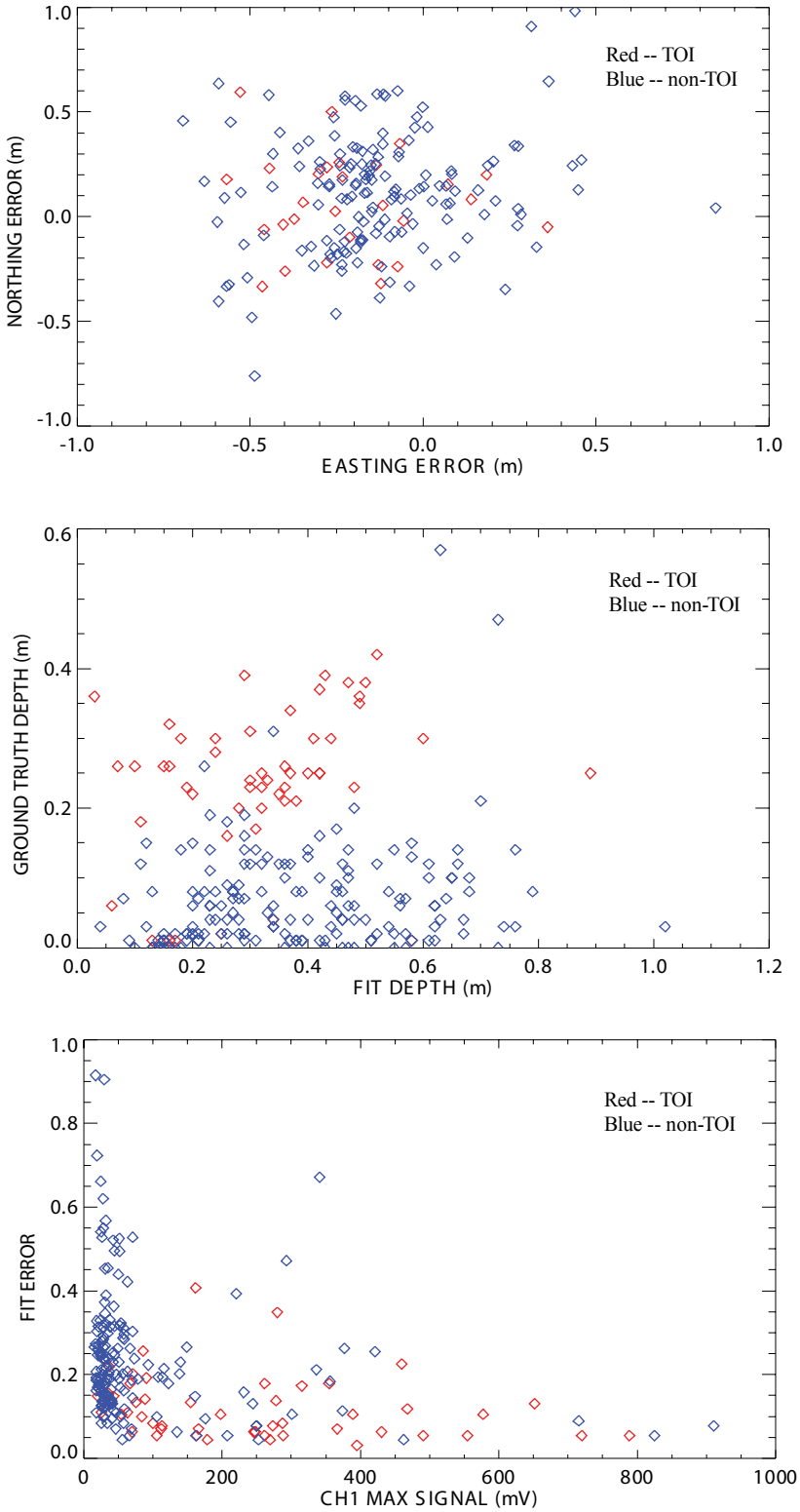


Figure 6-12. Plots showing scatter in the positioning, fitted versus ground truth depth, and fit error as a function of signal strength for the EM61 cart with no slope correction data.

EM61 CART – NO SLOPE CORRECTION APPLIED (2/4)

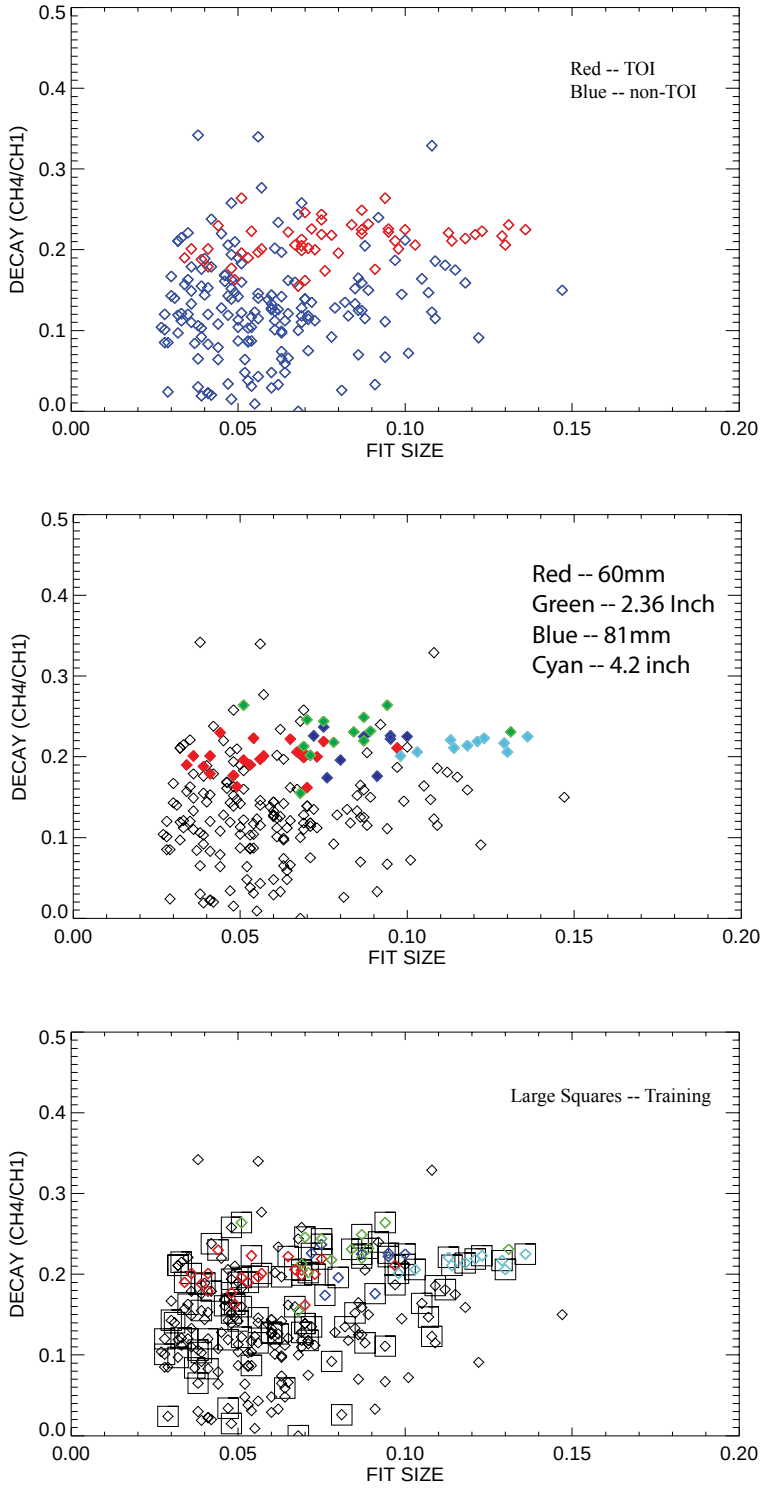


Figure 6-13. Plots of the fitted size versus decay. The three plots highlight different aspects of the training data for the EM61 cart with no slope correction.

EM61 CART – NO SLOPE CORRECTION APPLIED (3/4)

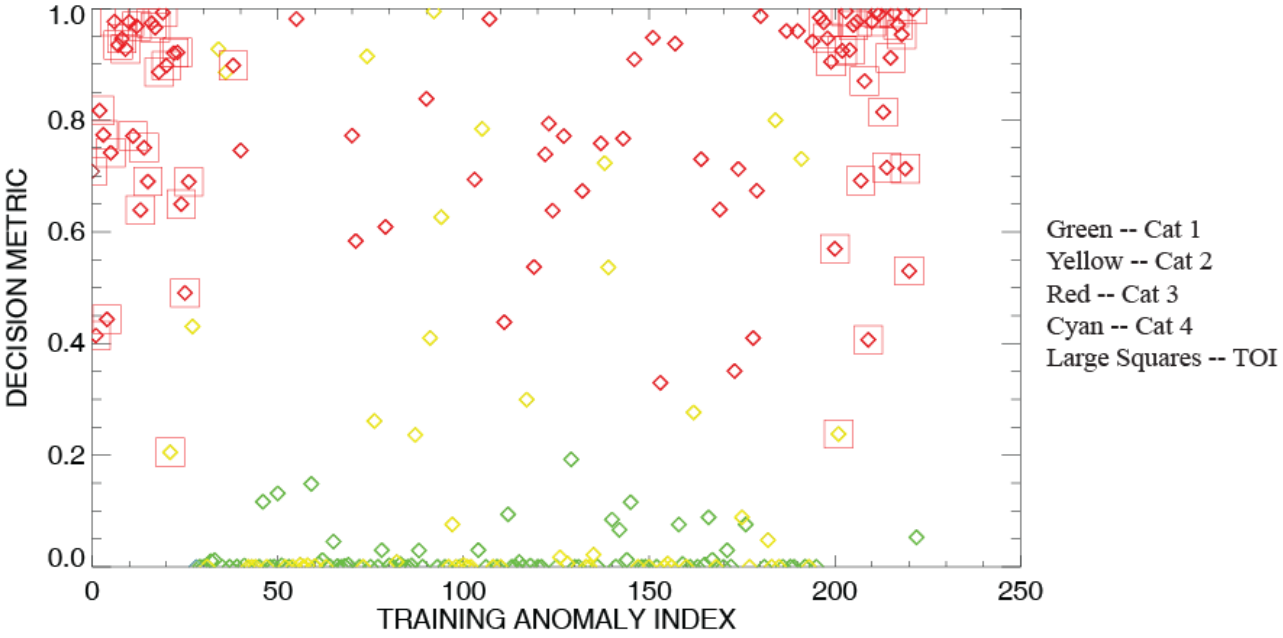


Figure 6-14. Plot of the decision metric for the EM61 cart with no slope correction data. Colors are used to indicate rank as indicated in the figure caption.

Thresholds used for drawing ROC curve (next page):

Threshold separating Category 1 and 2: 0.20

Threshold separating Category 2 and 3: 0.30

Table 6-2 Training Set Summary: Non slope corrected EM61 MK2 Cart

Category	Cultural	Munition Debris	No Contact	Soil	UXO
1	6	70	0	0	0
2	4	57	0	1	1
3	2	28	0	0	27
4	0	0	0	1	0
TOTAL	**12**	**155**	**0**	**2**	**28**

EM61 CART – NO SLOPE CORRECTION APPLIED (4/4)

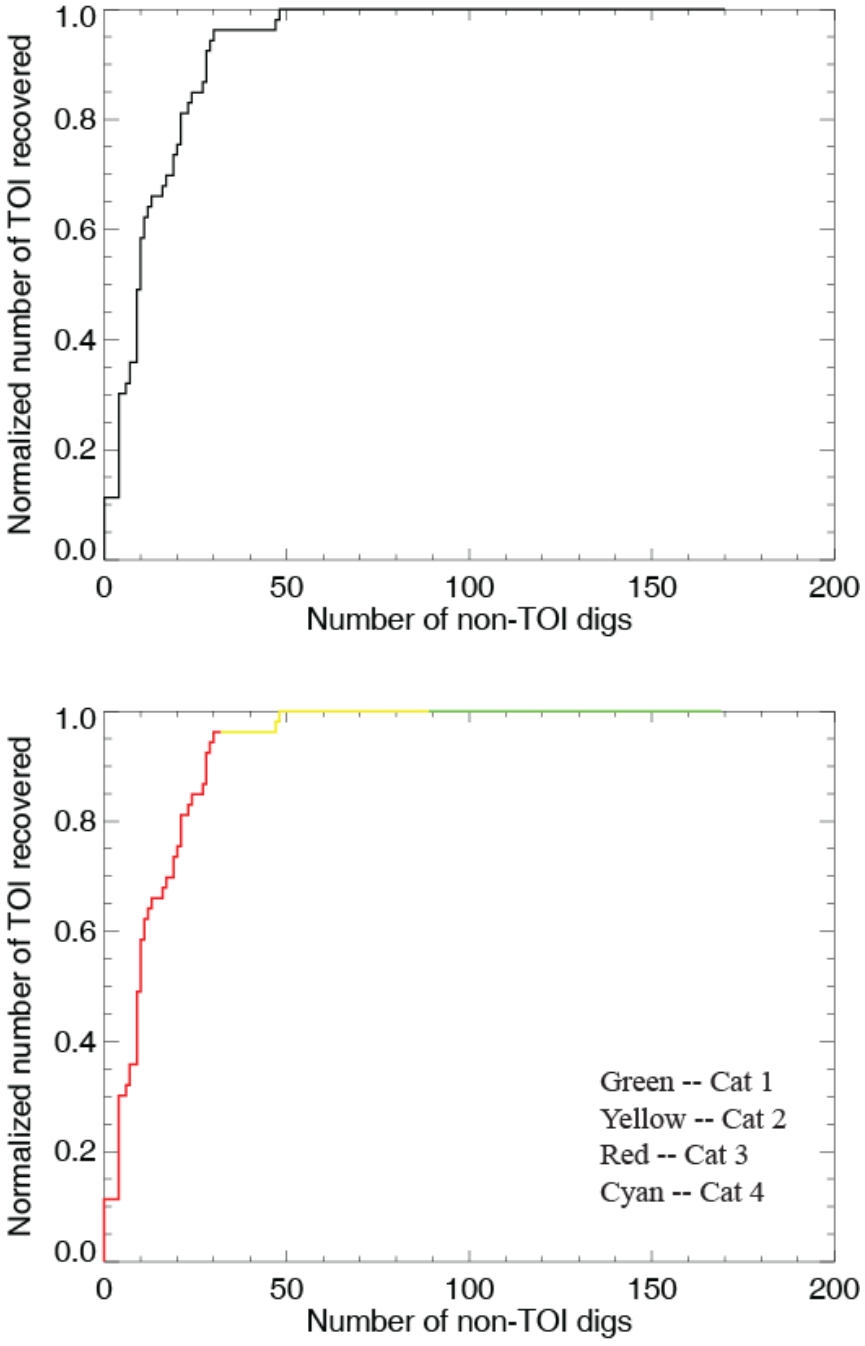


Figure 6-15. ROC curve of the EM61 cart with no slope correction training data. The top and bottom figures are identical except that the bottom graph is colored according to category.

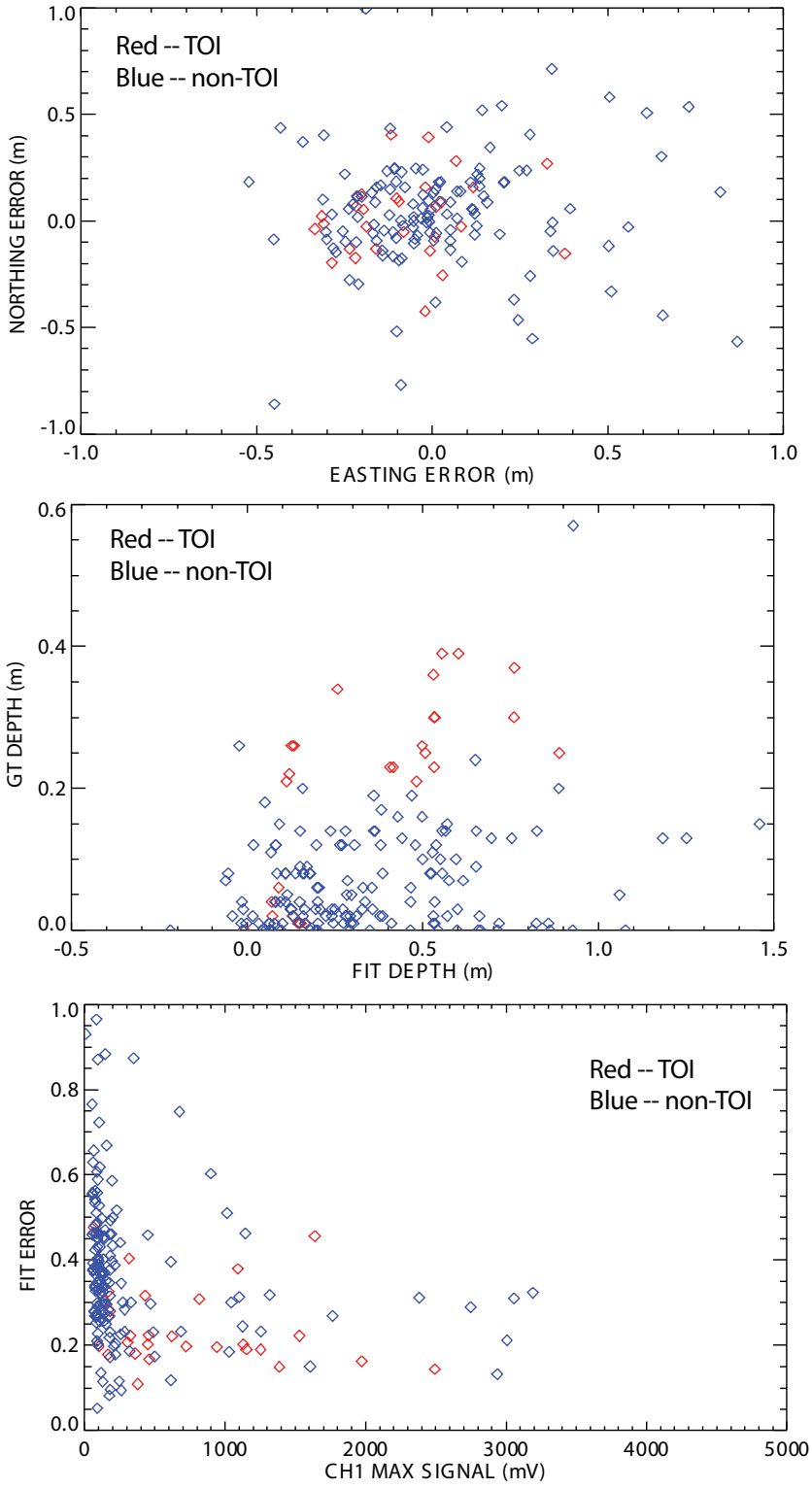


Figure 6-16. Plots showing scatter in the positioning, fitted versus ground truth depth, and fit error as a function of signal strength for the EM61 array data.

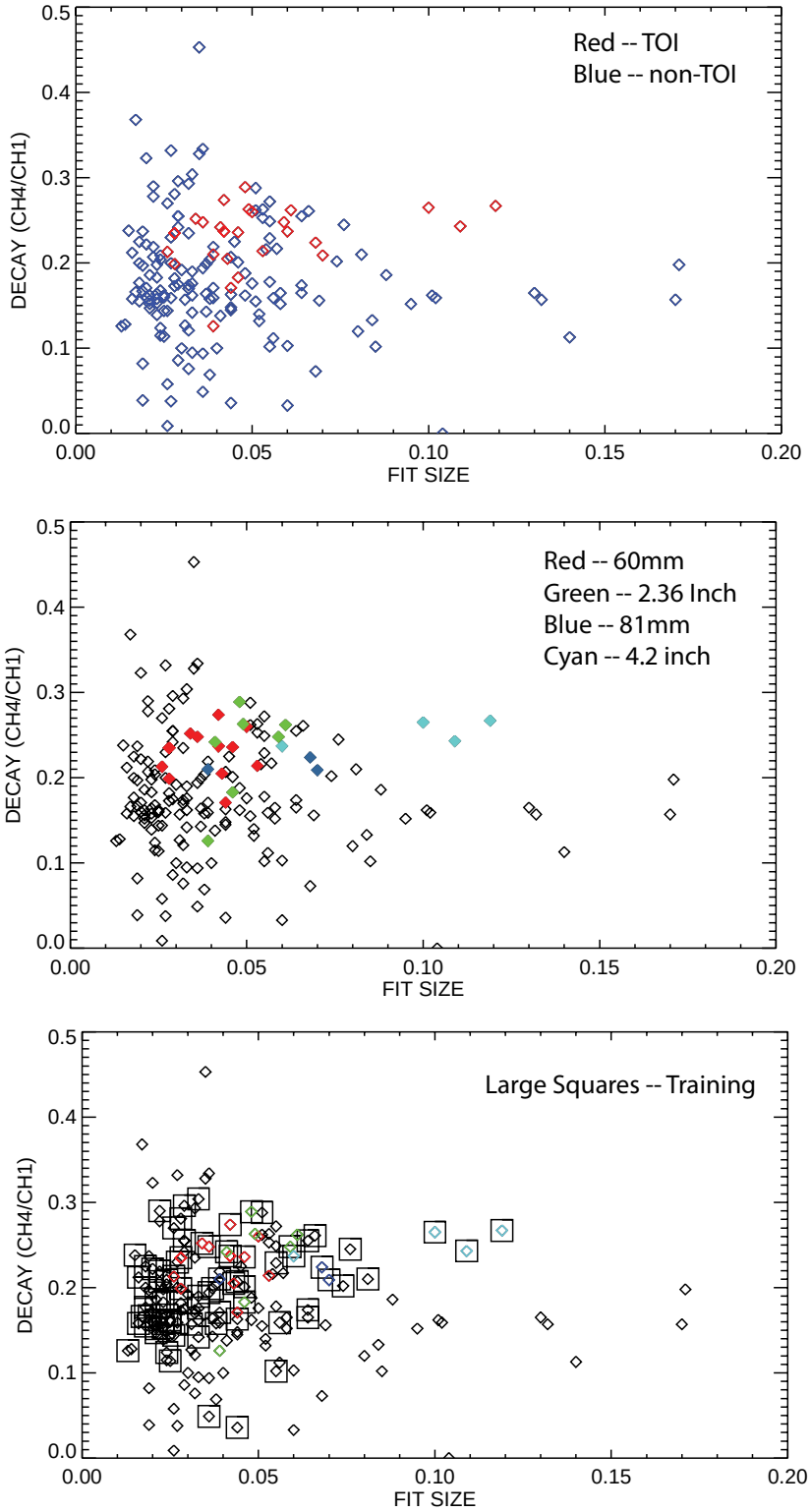


Figure 6-17. Plots of the fitted size versus decay. The three plots highlight different aspects of the training data for the EM61 array.

EM61 ARRAY (3/4)

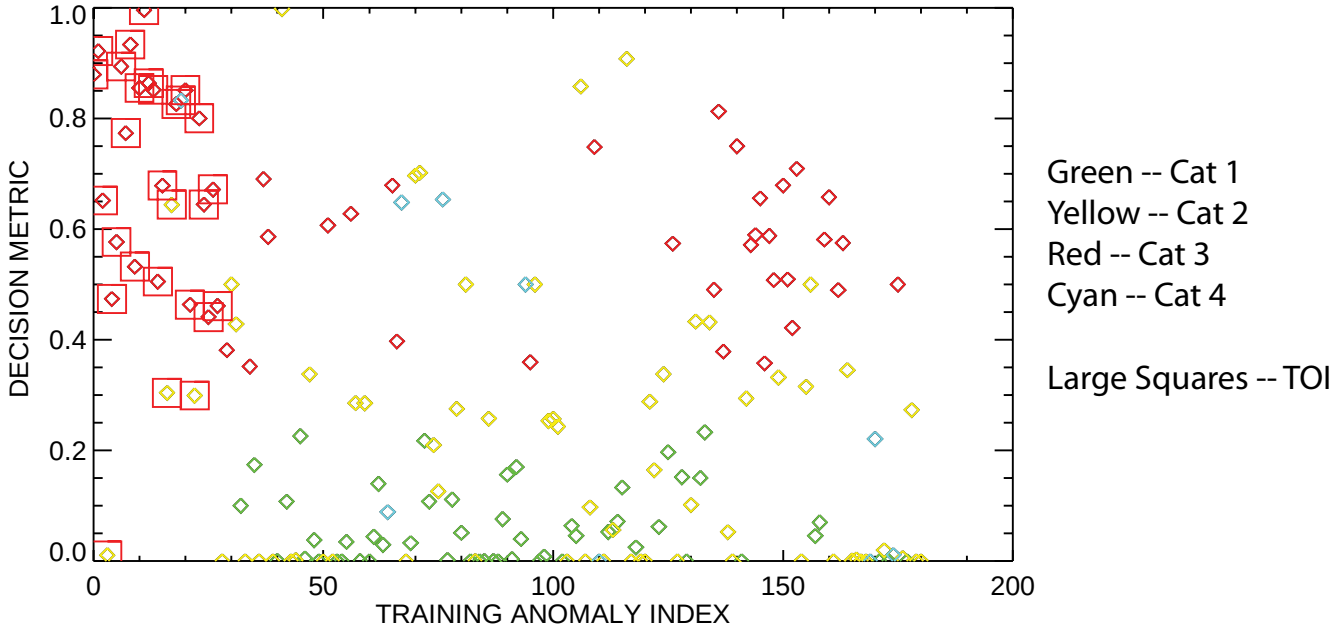


Figure 6-18. Plot of the decision metric for the EM61 array training data. Colors are used to indicate rank as indicated in the figure caption.

Thresholds used for drawing ROC curve (next page):

Threshold separating Category 1 and 2: 0.25

Threshold separating Category 2 and 3: 0.35

Table 6-3 Training Set Summary: EM61 Array

Category	Cultural	Munition Debris	No Contact	Soil	UXO
1	6	70	0	0	0
2	4	57	0	1	1
3	2	28	0	0	27
4	0	0	0	1	0
TOTAL	**12**	**155**	**0**	**2**	**28**

EM61 ARRAY (4/4)

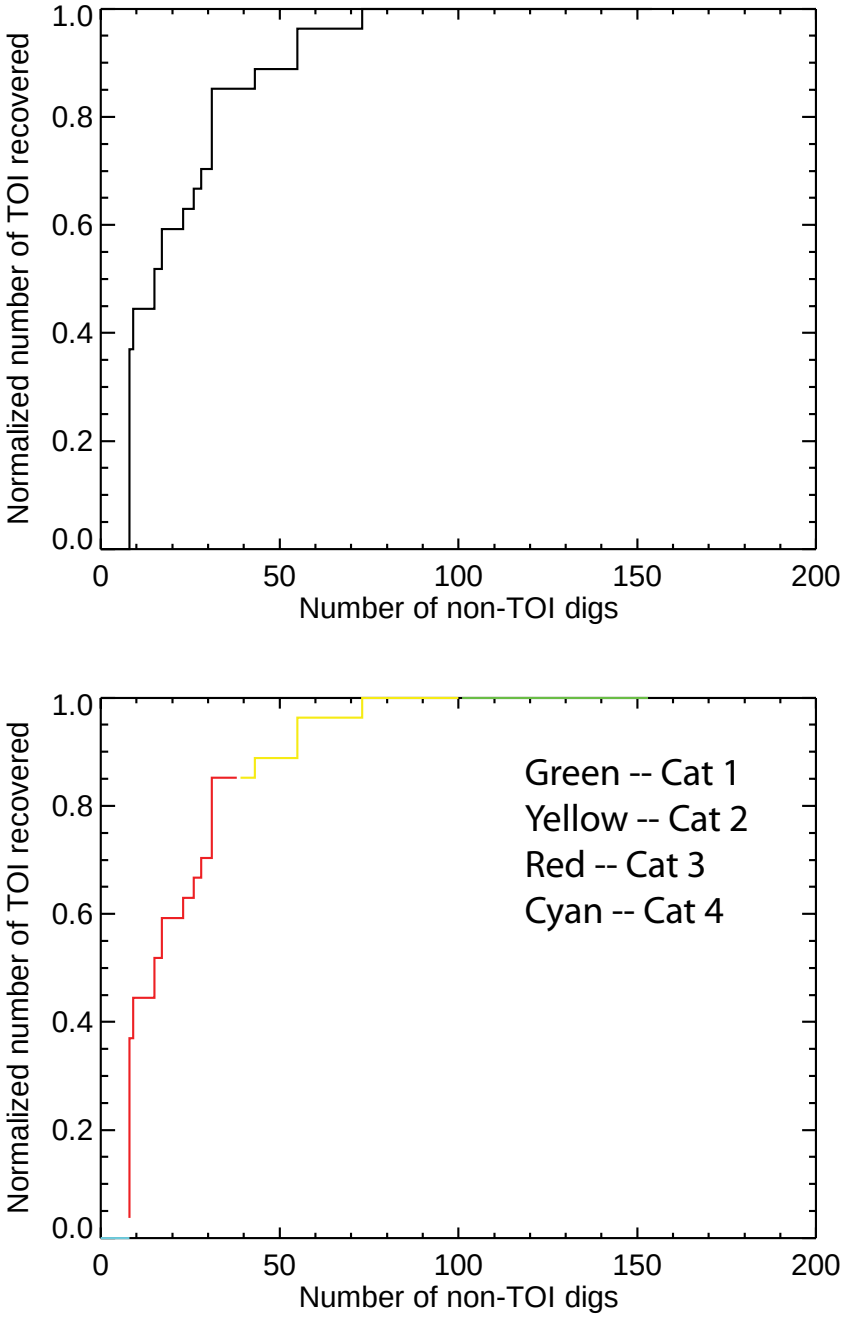


Figure 6-19. ROC curve of the EM61 array training data. The top and bottom figures are identical except that the bottom graph is colored according to category.

EM61 MSEMS (1/4)

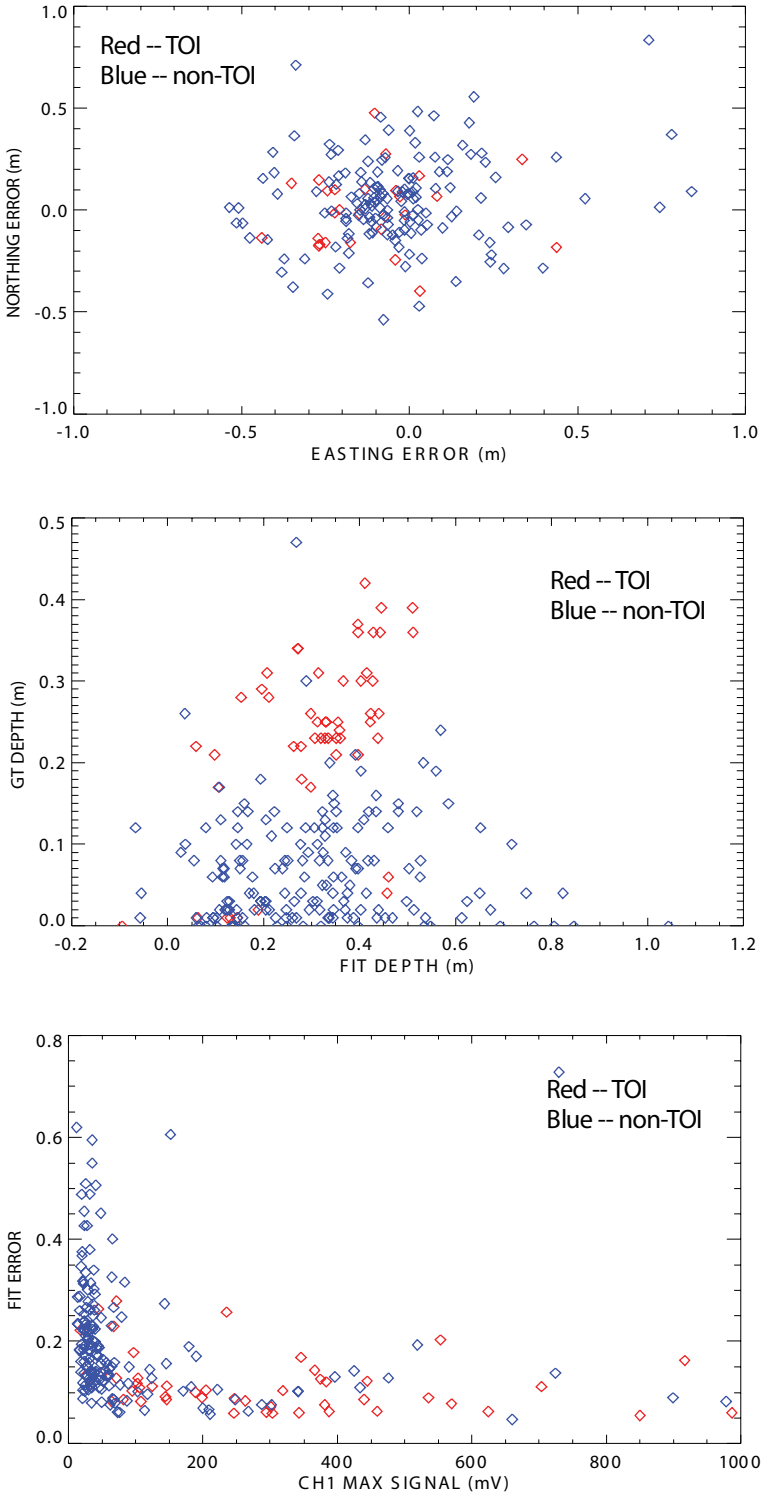


Figure 6-20. Plots showing scatter in the positioning, fitted versus ground truth depth, and fit error as a function of signal strength for the EM61 MSEMS data.

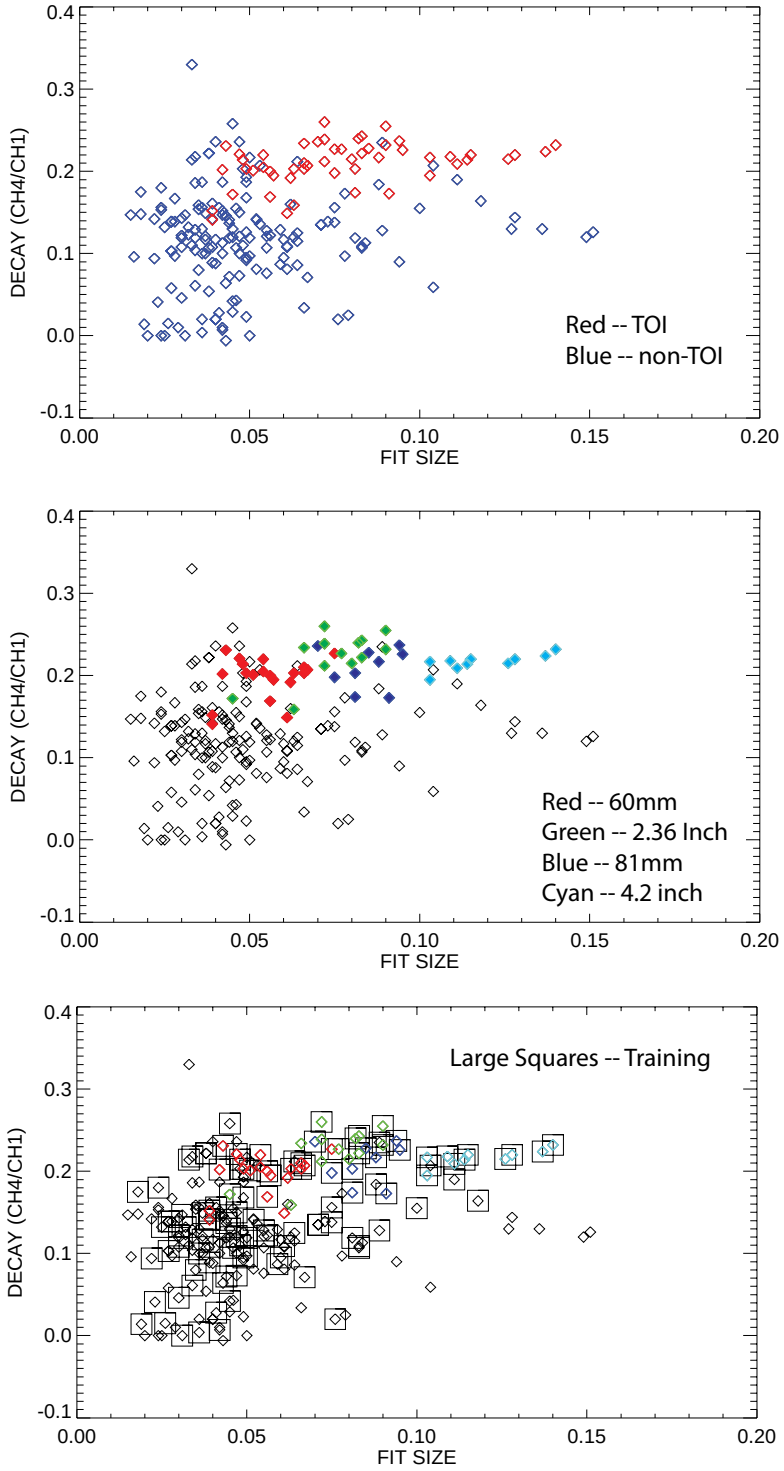


Figure 6-21. Plots of the fitted size versus decay. The three plots highlight different aspects of the training data for the EM61 MSEMS.

EM61 MSEMS (3/4)

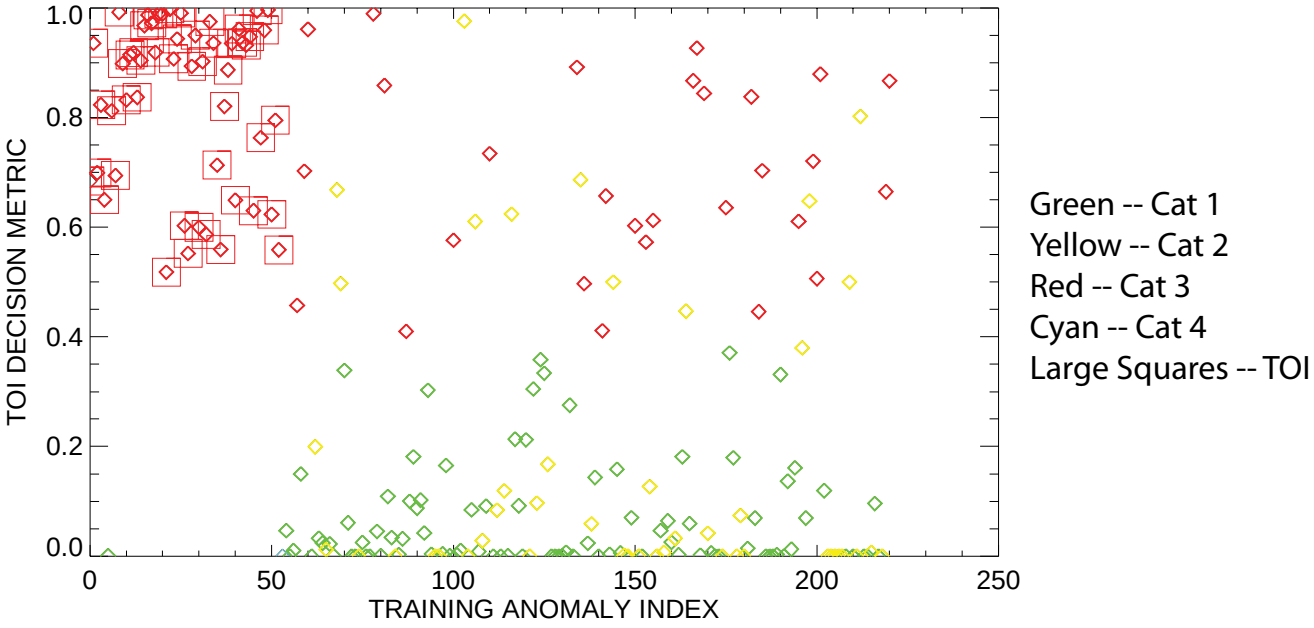


Figure 6-22. Plot of the decision metric for the EM61 MSEMS training data. Colors are used to indicate rank as indicated in the figure caption.

Thresholds used for drawing ROC curve (next page):

Threshold separating Category 1 and 2: 0.35

Threshold separating Category 2 and 3: 0.40

Table 6-4 Training Set Summary: EM61 MSEMS

Category	Cultural	Munition Debris	No Contact	Soil	UXO
1	6	70	0	0	0
2	4	57	0	1	1
3	2	28	0	0	27
4	0	0	0	1	0
TOTAL	**12**	**155**	**0**	**2**	**28**

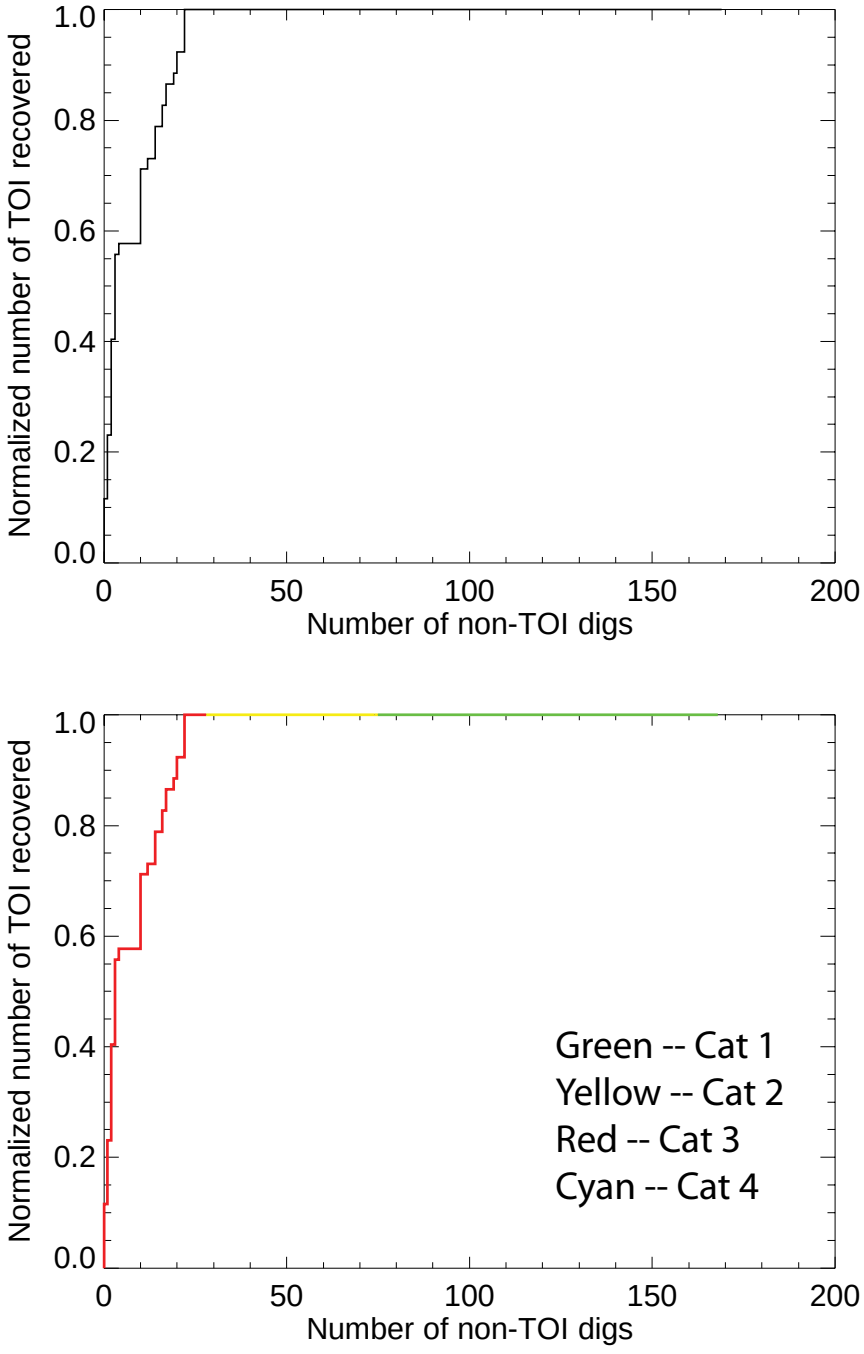


Figure 6-23. ROC curve of the EM61 MSEMS training data. The top and bottom figures are identical except that the bottom graph is colored according to category.

6.3.3 TEMTADS Data

The TEMTADS anomaly list was derived from the EM61 array survey. Experience with this array has shown that the first six or so timegates are affected by ringing from the transmitter. We therefore excluded these from consideration. Further, element 21 in the array was not working properly, and was also excluded from our analysis. In this section, we present scatter plots to provide a feel for the inverted target attributes.

Both our single- and multi-dipole models invert for the position, orientation, and principal axis polarizabilities (β's) as a function of timegate for the target(s). In Figure 6-24 to Figure 6-26, we show β plots for 3 targets from the Training Set. Figure 6-24 shows anomaly 473, a 4.2in mortar. As expected for axially symmetric ordnance, there is one larger β and two smaller, equal ones. In Figure 6-25, we plot anomaly 565, a 2.36in mortar. This munitions shows the greatest amount of variability of the 4 types known to be present at this site. In particular, the two smaller β's often show a greater separation than for the other 3 types. The behavior of this munitions item is believed to be a result of its composite nature. Finally, in Figure 6-26, we present the β plot for anomaly 5, a piece of 4.2in mortar frag. As is evident, the β's are quite distinct from those of the two ordnance items shown.

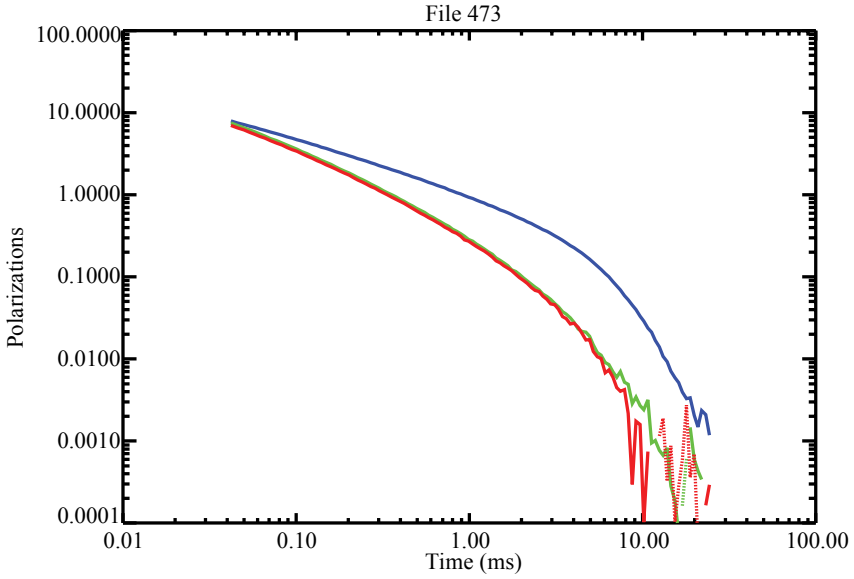


Figure 6-24. Principal axis polarization for a 4.2-inch mortar using TEMTADS data.

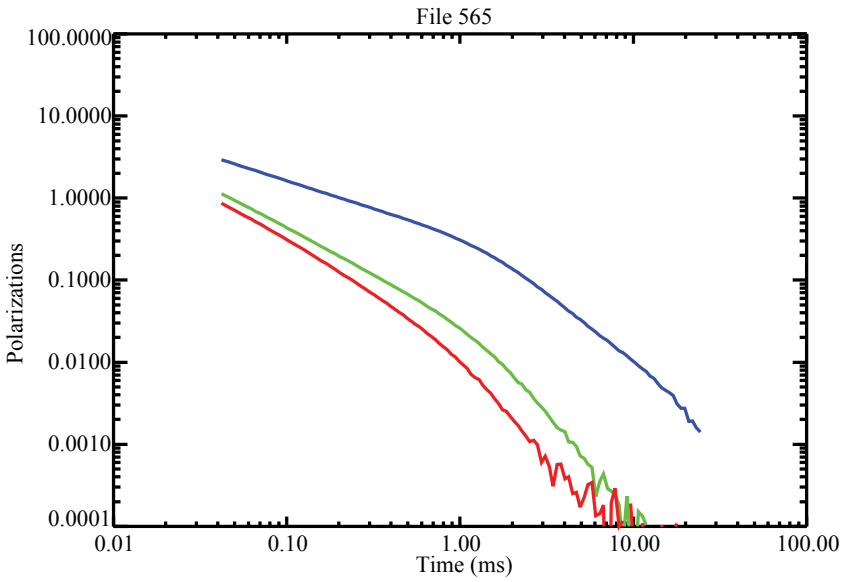


Figure 6-25. Principal axis polarization for a 2.36-inch mortar using TEMTADS data.

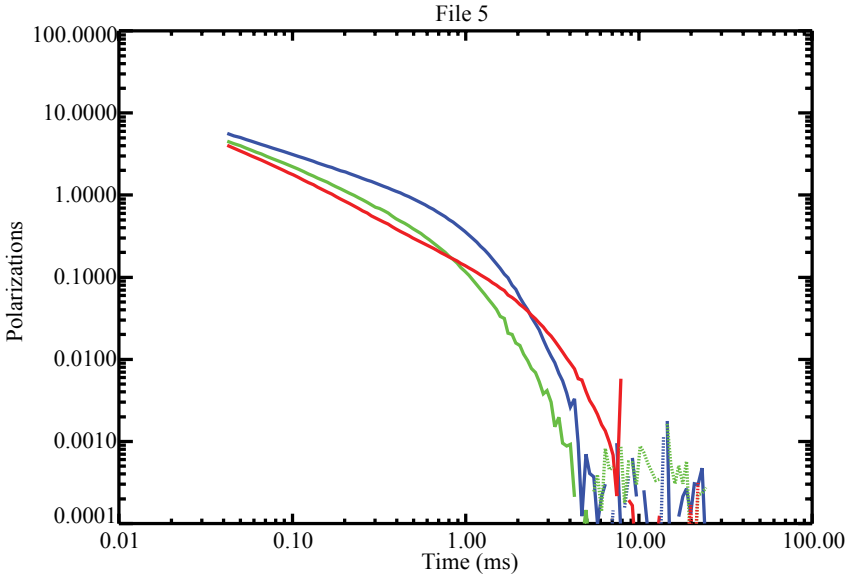


Figure 6-26. Principal axis polarization for a fragment of a 4.2-inch mortar using TEMTADS data.

Figure 6-27 plots the percentage fit error vs. the peak signal for all anomalies in the Training Set. For simplicity, the fit error depicted is only for single dipole fits. In this plot, as well as in Figure 6-28 and Figure 6-29, the four munitions types are shown as x's, with different colors distinguishing the four types. Clutter items are identified by blue squares. The symbol label 'miscellaneous TOI' identifies distinct UXO signatures from the site that were added to our library. The plot shows the expected inverse relations between fit error and signal strength.

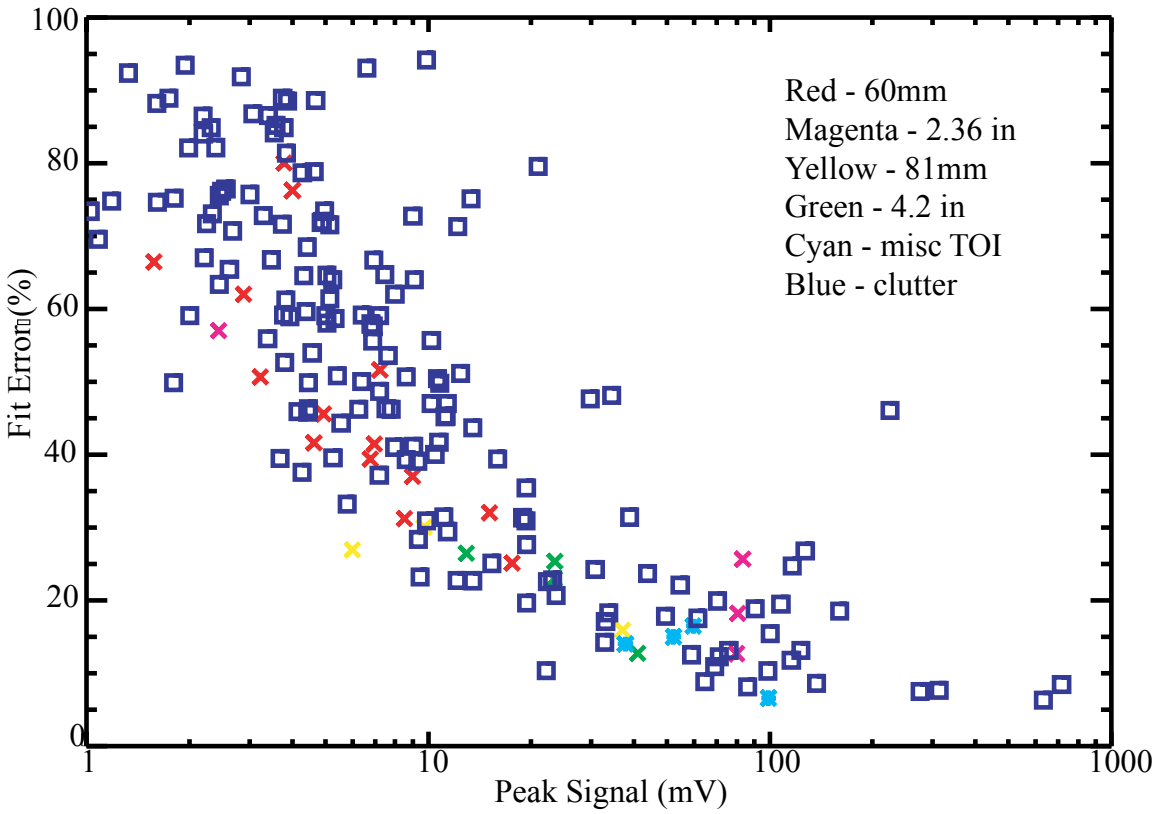


Figure 6-27. Fit Error vs. Peak Signal for the TEMTADS Training Set.

Classification for the TEMTADS dataset was primarily based on an algorithm which compared our derived polarizabilities with a library of known target signatures. In the majority of cases, these library signatures were taken under test stand conditions. The library contains a number of samples of the four ordnance types known to be present at the site, including measurements obtained at various target orientations, as well as several subtypes of these munitions. Whenever a measured signature appears distinct from those of the same type already present, we have added it as a separate entry.

Our primary library match algorithm compares the polarizabilities of an unknown target with each library entry based on 3 criteria: the amplitude of β_1, and the two shape parameters, β_2/β_1 and β_3/β_1. The difference in the values is computed at all timegates, excluding those where the values are negative. The differences were then plugged into a Gaussian. The "time constants" in these Gaussians were derived by examining the variability in the amplitude and shape parameters for a large number of objects for which ground truth was known. Medians were taken to avoid bad data points. Finally, the results from the 3 different criteria were averaged, producing a metric which ranges from 0 (worst possible fit) to 1 (perfect fit). Note that the procedure just described is not a library constrained match, i.e., we do not invert our data forcing the β's to be

those of each library object in turn, but rather simply compare our unconstrained polarizabilities to those of the library. As such, the comparison runs rapidly, and there is no need to reduce the number of separate types in the library to balance computation time.

For each anomaly, this match was performed on both the polarizabilities from the single dipole fit, and those of each target obtained by the multi-dipole fit. The metric was set to whichever set of β's produced the best match.

Due to the occasional "splitting" of the two smaller β's, as seen in Figure 6-25, we considered a second metric in which the 3rd comparison, namely the β_3/β_1 shape match, was dropped from the calculation. With Program Office approval, we submitted two separate dig lists, one using the 3 criteria match metric, and one using this 2 criteria match metric.

Prior to receiving the ground truth for the Training Set, several targets were observed which seemed to match the 60mm mortars in our library in shape, but were smaller in amplitude. These targets were later revealed to be 60mm's with the tail assembly missing. Since this constitutes a distinct signature, we added our best sample of these cases to our library. Similarly, it was necessary to add the non-munition items now classified as targets of interest to our library to ensure our ability to properly identify similar targets in the Test Set.

In Figure 6-28 and Figure 6-29, we plot our 3 criteria and 2 criteria metric, respectively, vs. peak signal for the Training Set. It can be seen that in both cases, the metrics produced a clear separation between clutter and ordnance/targets of interest. Two points should be noted. First, the 60mm (red x) which falls below the rest of the ordnance/targets of interest is of very low signal, and would be classified as "Cannot Analyze", as discussed in the following subsection. Second, since one 60mm and 3 of the targets of interest from the Training Set were added to the library (we had already added an example of a 2.36in rocket head), they produce a metric of 1, since they are matching to themselves.

In this subsection, we describe our decision rules for classifying anomalies into the 4 categories in the final dig list, as well as our criteria for determining overlapping signatures.

We classified as "Cannot Analyze" targets for which the inversion produced unphysical parameters, specifically, depths below a 2m cutoff which was based on the largest target expected on the site, and negative polarizabilities.

We classified targets as "Likely Munitions" based on the library match metrics previously described. The cutoff was determined to insure that all training targets revealed as UXO or TOI with moderate-to-good signal-to-noise were included. Our cutoff values were 0.8336 for the 3 criteria metric, and 0.8925 for the 2 criteria metric.

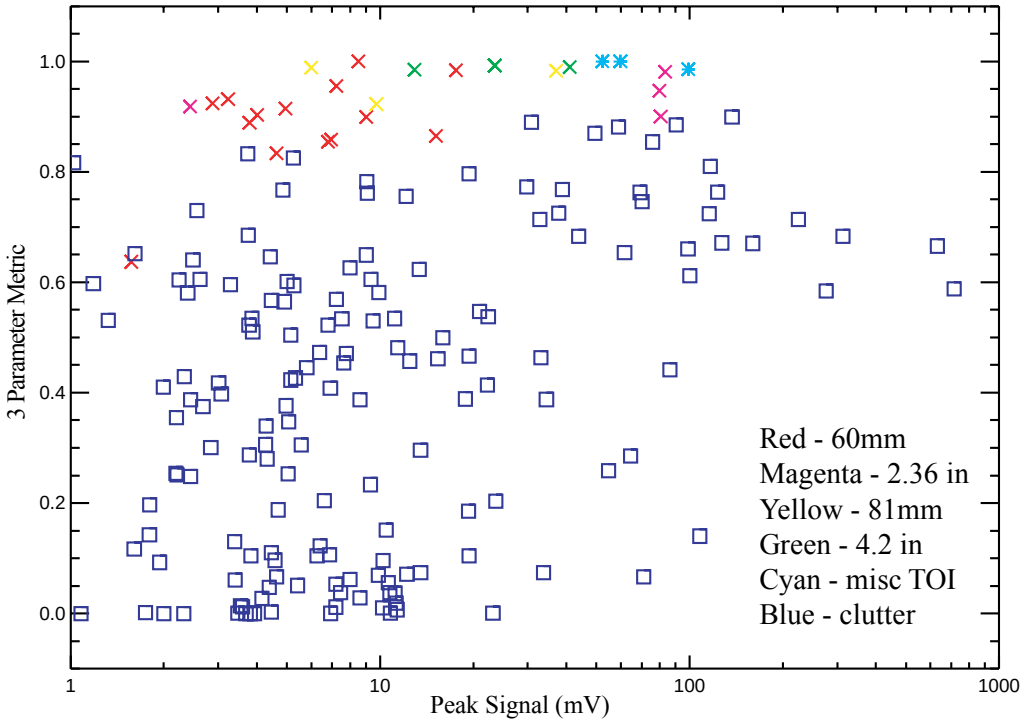


Figure 6-28. 3 Criteria Metric vs. Peak Signal for the TEMTADS Training Set.

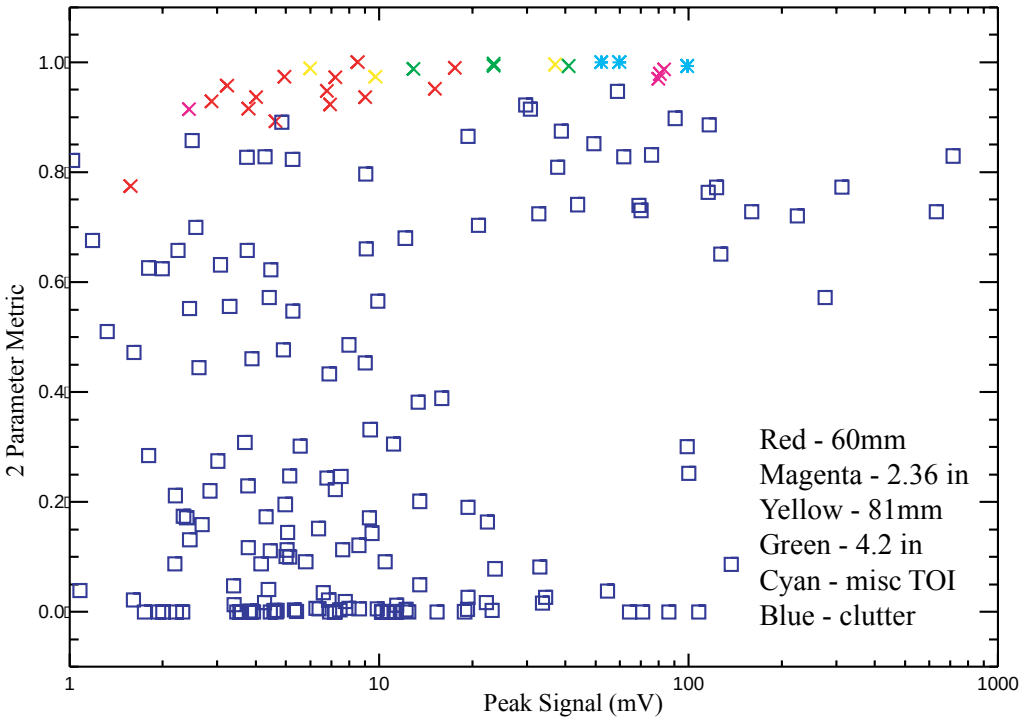


Figure 6-29. 2 Criteria Metric vs. Peak Signal for the TEMTADS Training Set.

Targets in the "Cannot Decide" category were comprised of four distinct types. The first type consisted of targets with very low signal-to-noise. Based on previous experiences with this instrument a value of 2mV/A was used. The second type was targets which suffer from serious overlap issues with neighboring anomalies, to the extent that we were not confident in our ability to extract meaningful features. The remaining two categories excluded cases from the first two. The third category consists of targets whose match metric fell within a "buffer zone" below the "Likely Munition" metric cutoff. The anomalies in this category possessed a match metric which is not sufficiently high to justify their being classified as "Likely Munitions", but which is not sufficiently low to classify them as "Likely Clutter". The actual values for this lower metric cutoff were determined from a visual inspection of matches in the Test Set, and were 0.7365 for the 3 criteria metric, and 0.8335 for the 2 criteria metric. Finally, the fourth type consisted of anomalies for which the metric fell below the "Likely Munition" cutoff, but which possess axial symmetry. This choice was made to take account of the fact that our library is finite. Exceptions to the axial symmetry rule were targets showing rapid decay (based on the most rapid decay for UXO in the complete library), and targets whose polarizabilities were of very small amplitude (based on the smallest amplitude UXO in the complete library). We conservatively defined axial symmetry as a median agreement of 50% between the two smaller betas.

We placed targets in the "Likely Clutter" category whose metric fell below that of the second "buffer zone" cutoff and which did not show axial symmetry as previously defined.

The "Likely Clutter" category was divided into two: low signal-to-noise and moderate to high signal-to-noise. Within each of these subcategories, ranking were by the library match metric, such that the lower the metric, the higher the rank. However, all anomalies in the low SNR subcategory were ranked lower than those in the moderate to high SNR subcategory. The rationale behind this was that low SNR can decrease the quality of the match metric, and the targets for which we have the highest confidence of being clutter are those with a low metric and decent SNR. Similarly, the "Cannot Decide" category was divided into three based on four types of entries, with the targets having low signal and overlap issues having the highest rank, followed by those with moderate-to-high signal and no serious overlap issues, with those targets having a metric above the "buffer zone" cutoff being giving a lower ranking (less likely to be clutter) than those below. The ranking within each subcategory were by the metric. Finally, the "Likely Munition" category was ranked solely by the metric, with the largest rank going to the largest metric value.

Since it is not obvious how to incorporate either the axial symmetry parameter or the low SNR parameter into our probabilities, we set the probability equal to the 1 minus the metric value for all targets. Unfortunately, with the exception of the "Likely Munition" category, this means that there was not in general a 1 to 1 mapping between the ranks and the probabilities.

Contour plots were made of each anomaly both as a QC check and as a means of determining whether the anomaly suffers from overlapping signatures. An entry was made in a spreadsheet giving some indication of the extent of the overlap, if any. However, in some cases, this overlap may be from multiple targets within the anomaly, as opposed to being caused by a separate

anomaly. Therefore, we used this entry in conjunction with the knowledge of the location of other anomalies relative to the target of interest. The edge of the TEMTADS is 1m from a target located under the array center. Thus, we declared any anomaly for which the spreadsheet entry indicates overlap, and for which at least one target is within 1.5m of the anomaly, as an overlapping signal. In addition to this baseline case, we have batch fit all the TEMTADS anomalies using SAICs multi-target solver as described in section 6.4.

In Figure 6-30 and Figure 6-31, we plot the library match metric for the 3 and 2 criteria matches vs. fiducial number for all the Training Set anomalies. In these figures, we matched the color scheme for the sample dig lists, by coloring all "Likely Clutter" green, all "Cannot Decide" yellow, and all "Likely Munition" red. For readability, the "Cannot Analyze" targets are colored cyan. We have included the two cutoffs discussed above as horizontal lines. Corresponding ROC curves are shown in Figure 6-32 and Figure 6-33.

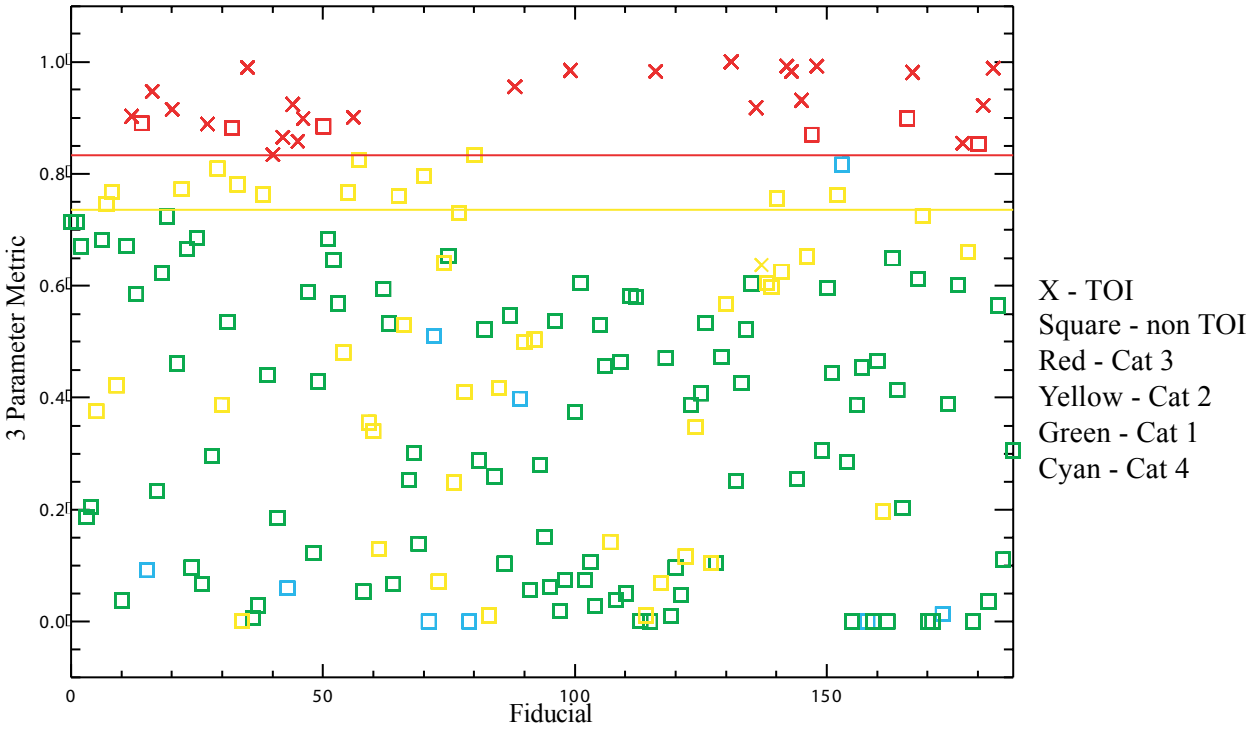


Figure 6-30. Category display of 3 Criteria Metric vs. Fiducial number for the TEMTADS Training Set.

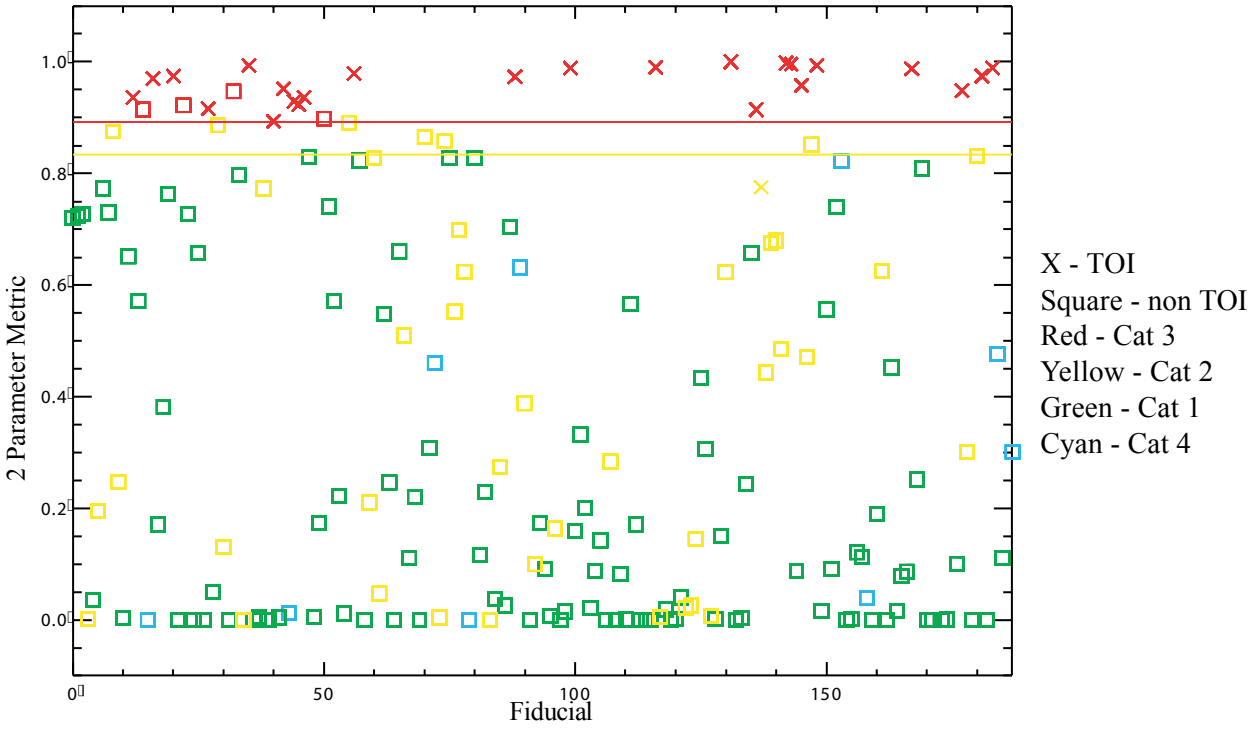


Figure 6-31. Category display of 2 Criteria Metric vs. Fiducial number for the TEMTADS Training Set.

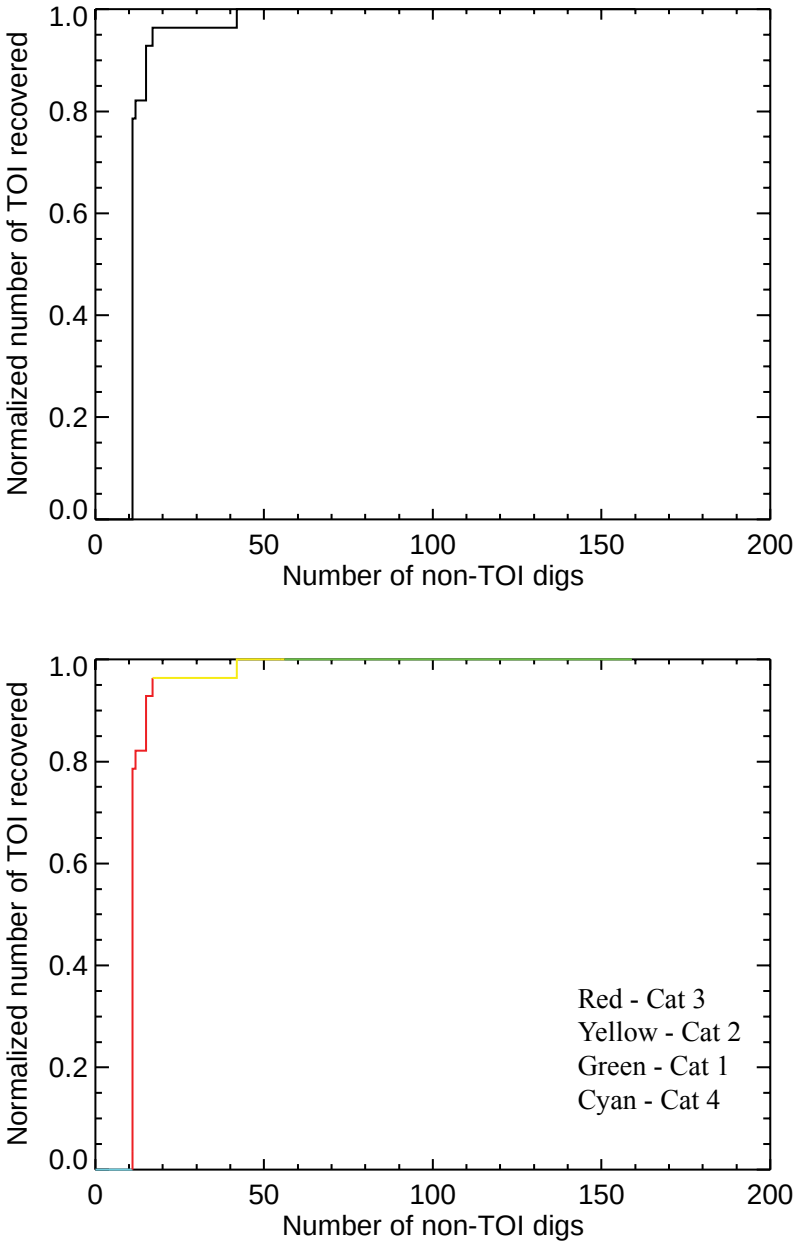


Figure 6-32. TEMTADS ROC curve for the 3 criteria metric (top in black and white; bottom is color coded to indicate Category rankings).

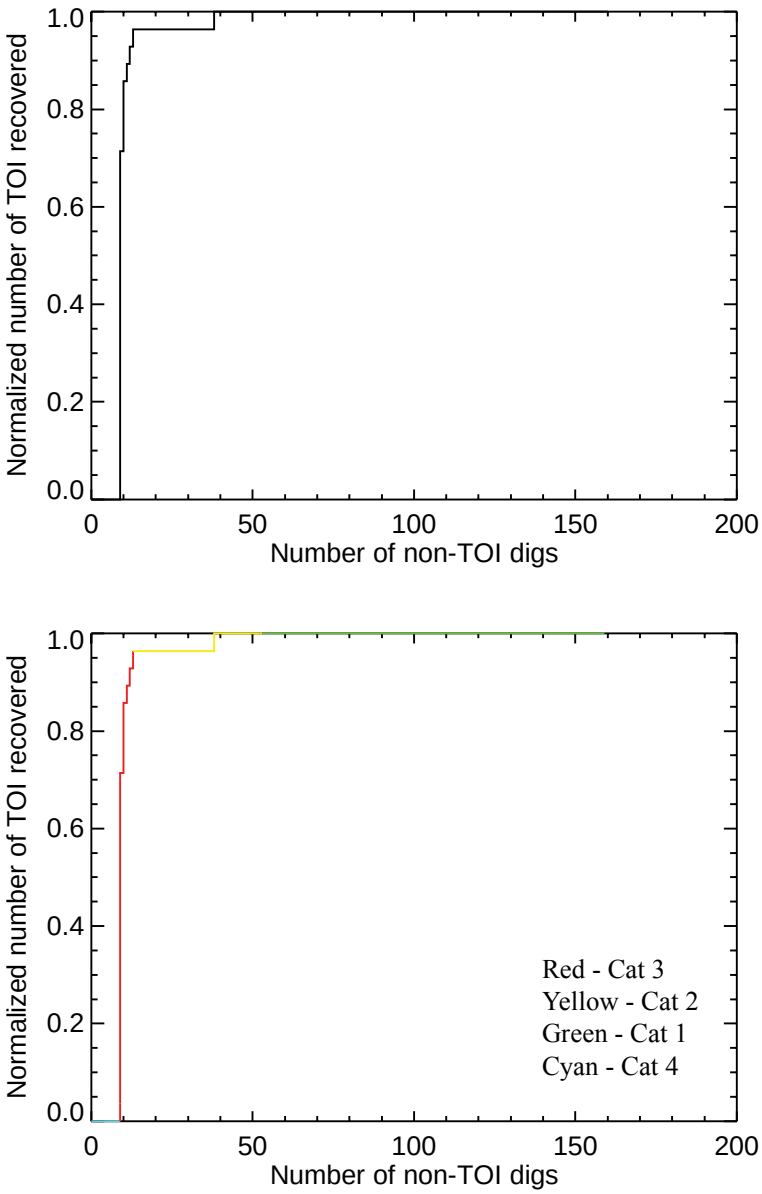


Figure 6-33. TEMTADS ROC curve for the 2 criteria metric (top in black and white; bottom is color coded to indicate Category rankings).

6.3.4 Metal Mapper Data

The Metal Mapper cued-mode anomaly list was derived from the dynamic Metal Mapper survey. The original list contained 2176 targets. QC analysis conducted while the cued survey was still in progress revealed a non-negligible offset between sensor and anomaly location for a number of targets. As a result, 307 targets were re-measured. Both the original and the repeat measurements are included in the data. A small number of extra targets which had not been initially chosen from the dynamic survey were also added. Additionally, subsequent to leaving the site, it was discovered that the data file for anomaly 1483 was corrupted. This results in 2491 total measurements. After matching the original and repeated measurements and determining cases where two of more original measurements were for the same anomaly, a number of targets were removed from the original list. The final list now contains 1560 targets. Ground truth for 153 of these targets was supplied for training purposes.

In Figure 6-34 to Figure 6-36, we show β plots for three targets from the training set. Figure 6-34 shows anomaly 2267, a 4.2in mortar. In Figure 6-35, we plot anomaly 3, a 2.36in mortar. Finally, in Figure 6-36, we present the β plot for anomaly 481, a piece of 4.2in mortar frag. Similar to the TEMTADS data the 4.2 in mortar shows good axial symmetry while the 2.36in mortar shows higher variability and separation in the two smaller β's and the frag produces β's that are quite distinct from those of the two ordnance items shown.

Figure 6-37 plots the percentage fit error vs. the peak signal for all anomalies in the training set. For simplicity, the fit error depicted is only for the single dipole fits. In this plot, as well as in Figure 6-38 and Figure 6-39, the 4 munitions types are shown as x's, with different colors distinguishing the four types. The four non-munition targets of interest are depicted with cyan asterisks, and clutter are represented by blue squares. The plot shows the expected behavior, with the general trend of fit error increasing for decreasing signal. We suspect that the slight upturn at the highest signals may be the result of saturation or near-saturation effects.

Classification for the Metal Mapper dataset was primarily based on an algorithm which compares our derived polarizabilities with a library of known target signatures obtained with the TEMTADS. Both the Metal Mapper and the TEMTADS were developed at G & G Sciences, and have essentially the same waveform. Experience with the precursor of the Metal Mapper, the AOL2, demonstrated that the polarizabilities derived from that system could be well-matched to ones derived from TEMTADS data by the application of a time lag and scale factor to the former. We used Metal Mapper data taken in the Calibration Strip to determine the appropriate lag and scaling correction to match the Metal Mapper to the TEMTADS.

For each Metal Mapper anomaly in the training and test sets, we followed the following steps prior to the library matching: First, the time lag derived as described above was applied to the Metal Mapper timegates. Second, the derived scale factor was applied to the Metal Mapper polarizabilities. Third, any TEMTADS timegates occurring later or earlier than those of the Metal Mapper were discarded. Finally, the Metal Mapper polarizabilities were interpolated to the TEMTADS timegates.

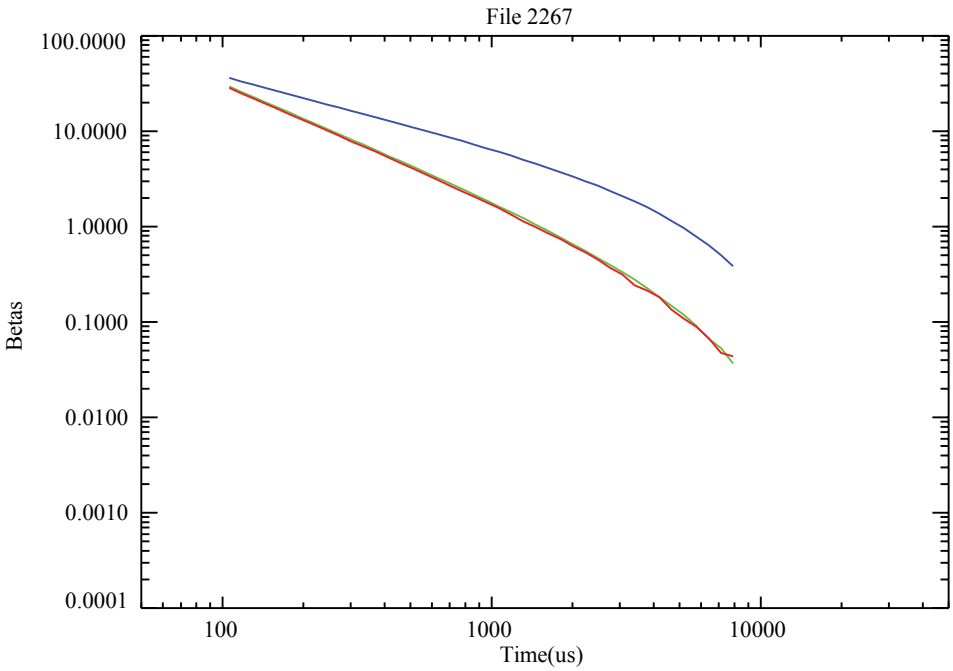


Figure 6-34. Principal axis polarizabilities (Beta's) for Metal Mapper target id 2267 a 4.2-inch mortar.

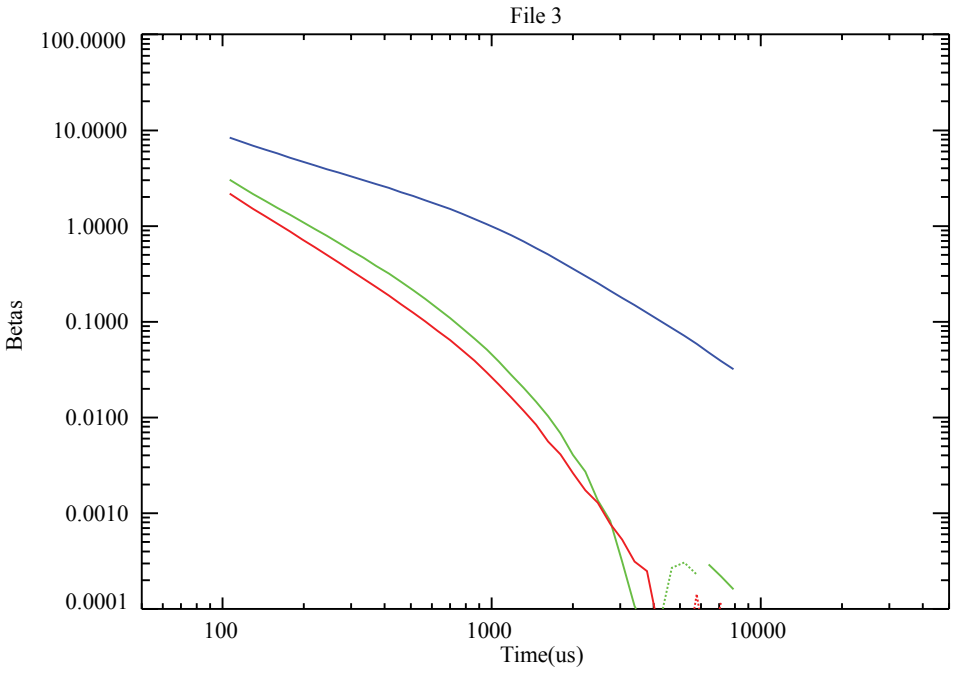


Figure 6-35. Principal axis polarizabilities (Beta's) for Metal Mapper target id 3 a 2.36-inch mortar.

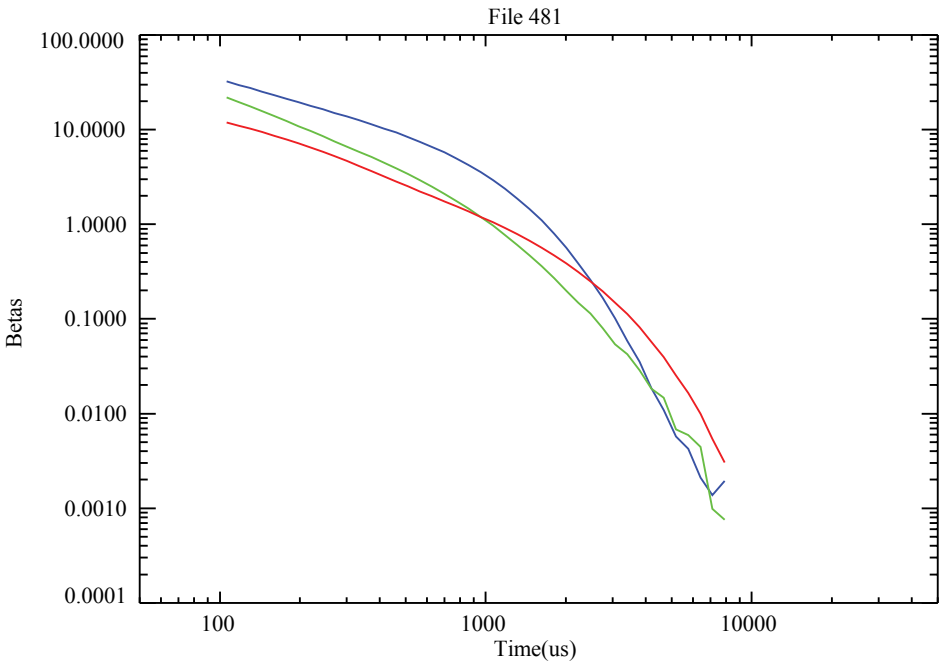


Figure 6-36. Principal axis polarizabilities (Beta's) for Metal Mapper target id 481 a 4.2-inch mortar fragment.

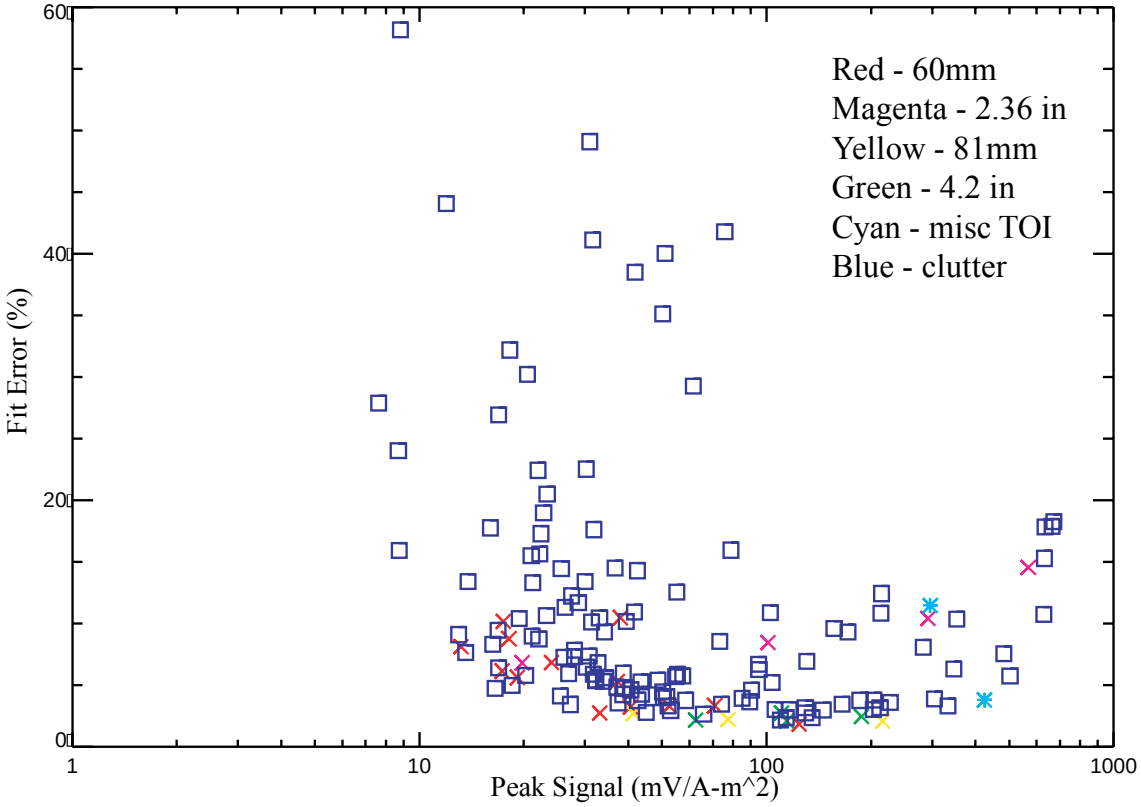


Figure 6-37. Fit Error vs. Peak Signal for the Metal Mapper Training Set.

In the majority of cases, the signatures in the library were taken under test stand conditions. The library contains a number of samples of the 4 ordnance types known to be present at the site, including measurements obtained at various target orientations, as well as several subtypes of these munitions. Whenever a measured signature appears distinct from those of the same type already present, we have added it as a separate entry.

The primary library match algorithm used was the same used for the TEMTADS which was described in the previous section. We also submitted two separate dig lists, one using the 3 criteria match metric, and one using this 2 criteria match metric.

In Figure 6-38 and Figure 6-39, we plot our 3 criteria and 2 criteria metric, respectively, vs. peak signal for the training set. It can be seen that in both cases, the metrics produce a clear separation between clutter and ordnance/targets of interest.

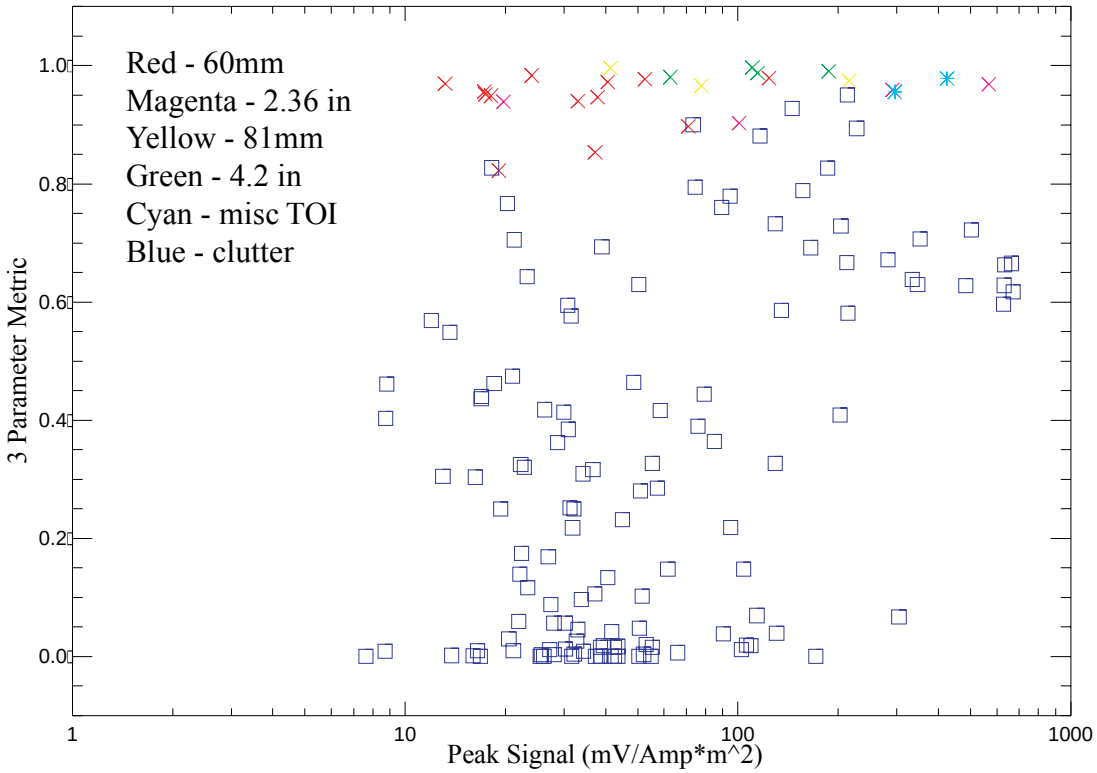


Figure 6-38. 3 Criteria Metric vs. Peak Signal for the Metal Mapper Training Set.

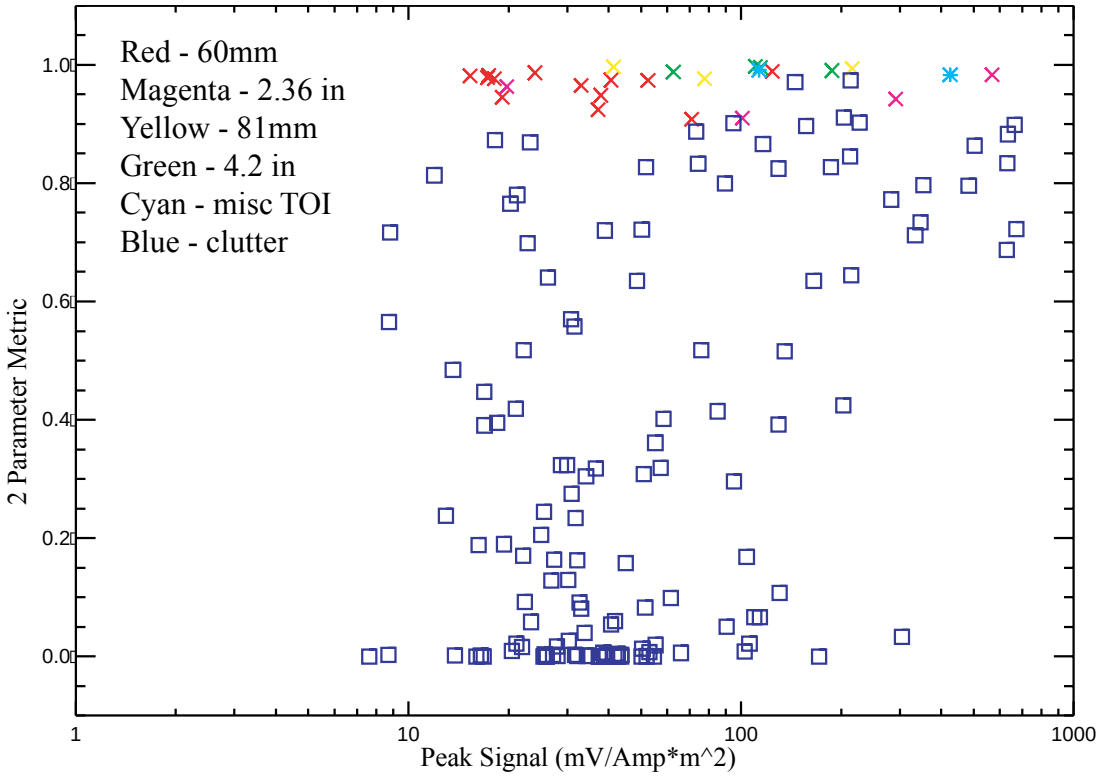


Figure 6-39. 2 Criteria Metric vs. Peak Signal for the Metal Mapper Training Set.

The decision rules for classifying the Metal Mapper anomalies were the same as those described for the TEMTADS except different thresholds were used and the "Cannot Decide" category only consisted of 2 distinct types. For the "Likely Munitions" category the cutoff values were 0.8224 for the 3 criteria metric, and 0.8987 for the 2 criteria metric.

The first type of targets in the "Cannot Decide" category were ones which suffered from serious overlap issues with neighboring anomalies, to the extent that we were not confident in our ability to extract meaningful features. The second type was those targets not included in the first type for which the metric falls below the "Likely Munition" cutoff, but which possess axial symmetry. We note that the signal to noise for this dataset is sufficiently good that we will not include any targets in this category based on that criterion.

The targets were ranked in a similar fashion to the TEMTADS ranking. Targets within each of the three "Can Analyze" categories were ranked solely by their metric value, with the largest rank going to the largest metric value. The "Cannot Decide" category was divided into two based on the two types of entries as described above. Targets having overlap issues were given the highest rank, followed by targets possessing axial symmetry. Again, the ranking within each of the two subgroups were based solely on the metric.

Although the Metal Mapper receiver cubes are spread out in various locations, the design of the system does not allow any obvious means of obtaining a contour plot for a given anomaly analogous to systems with single-plane transmitters and receivers. Thus there is no visual clue as to whether overlap occurs. Therefore, we simply used the knowledge of the location of other anomalies relative to the target of interest. The furthest receiver cubes are only 40cm from the center of the sensor, well within the 60cm used to declare whether anomalies should be considered separate or not. Thus, we declared any anomaly for which at least one target was within 1.5m of the anomaly, as an overlapping signal. In addition to this baseline case, we batch fit all Metal Mapper anomalies using SAICs multi-target solver as described in Section 6.4.

In Figure 6-40 and Figure 6-41, we plot the library match metric for the 3 and 2 criteria matches vs. anomaly number for all the Training Set anomalies. In these figures, we have attempted to match the color scheme for the sample dig lists, by coloring all "Likely Clutter" green, all "Cannot Decide" yellow, and all "Likely Munition" red. For readability, the "Cannot Analyze" targets are colored cyan. We have included the cutoff discussed above as a horizontal line. We used our ranking system described above to produce the ROC curves shown in Figure 6-42 and Figure 6-43.

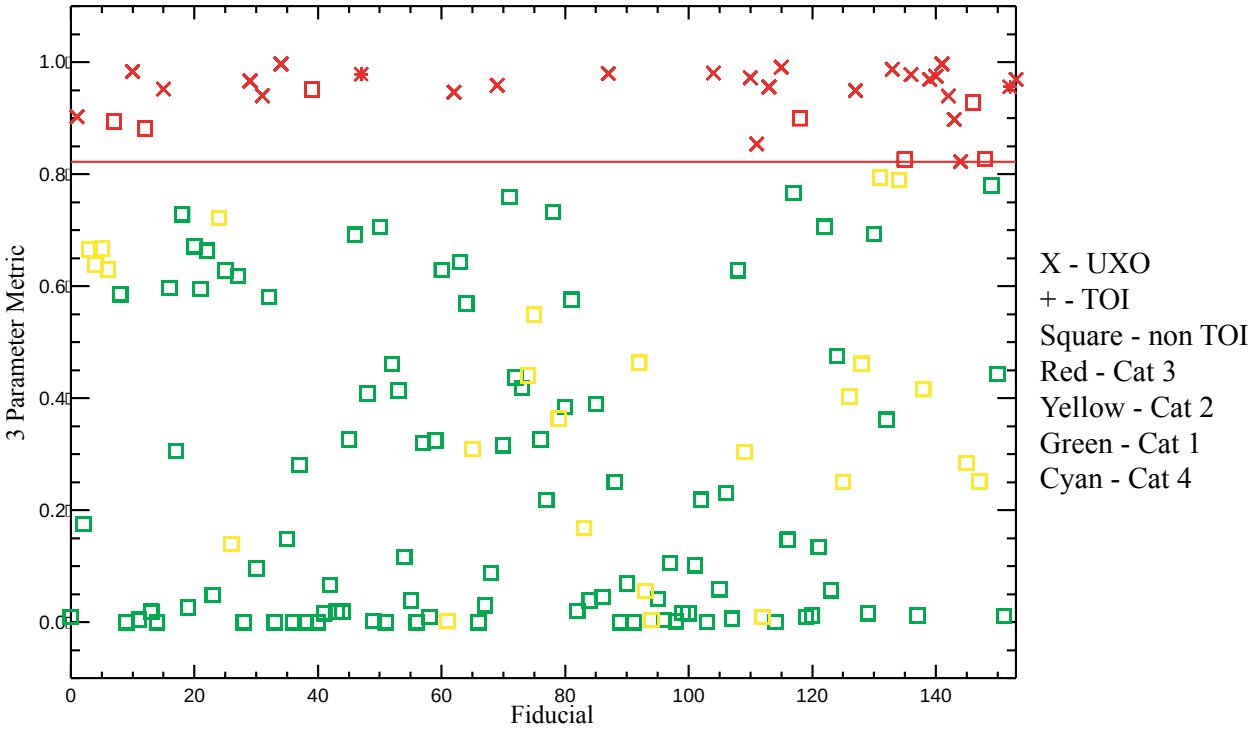


Figure 6-40. 3 Criteria Metric vs. Fiducial for the Metal Mapper Training Set, Rank-Coded

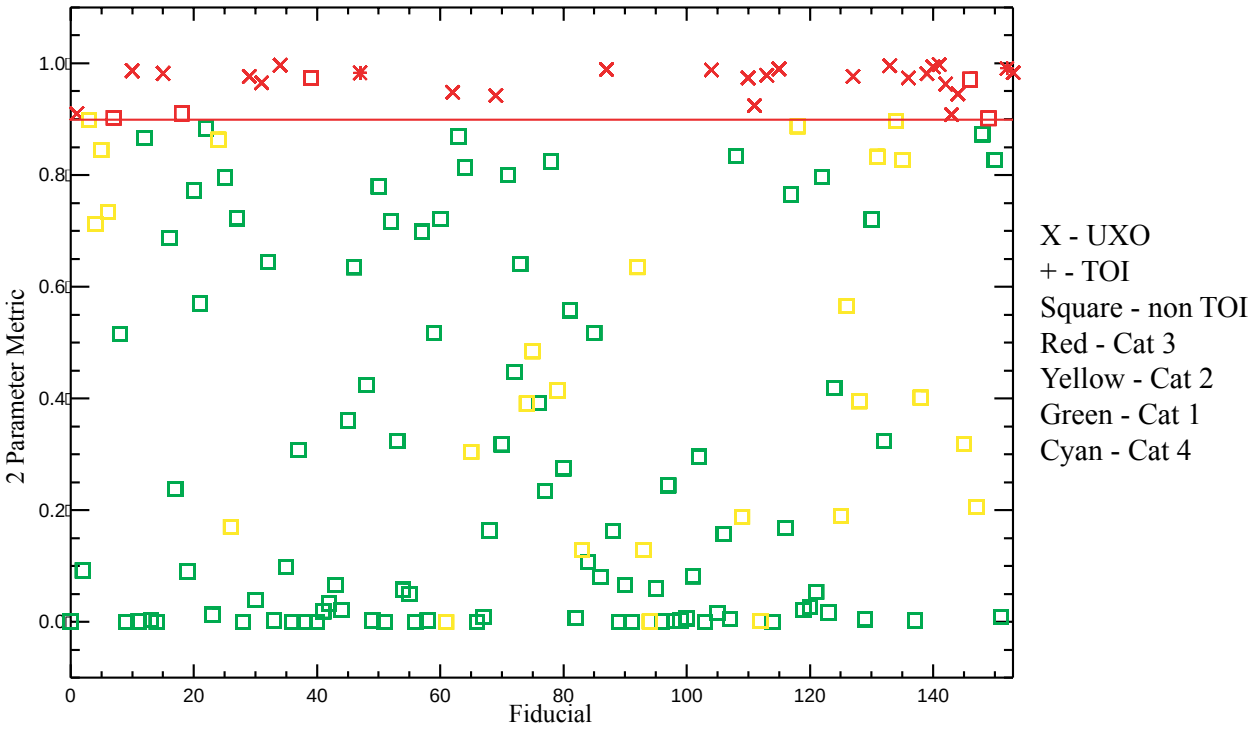


Figure 6-41. 2 Criteria Metric vs. Fiducial for the Metal Mapper Training Set, Rank-Coded.

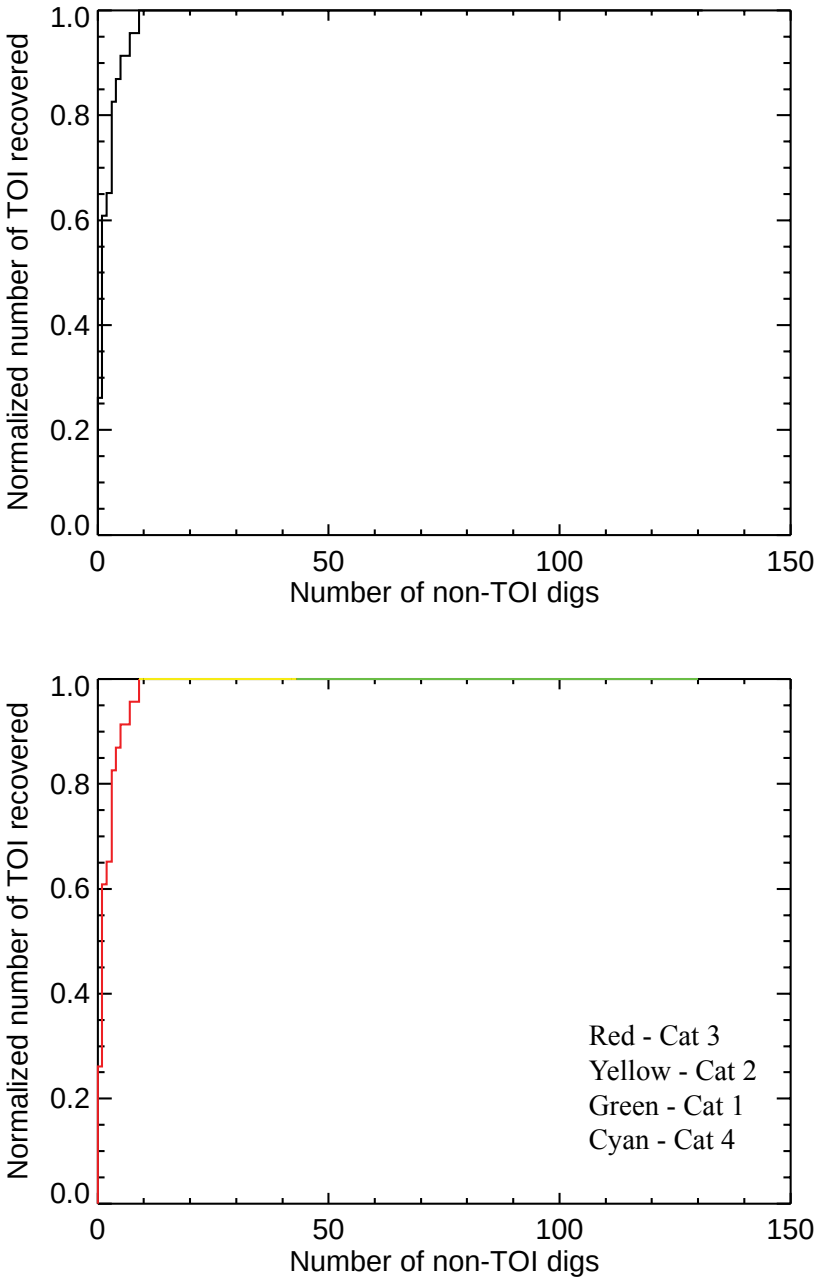


Figure 6-42. Metal Mapper ROC curve for the 3 Criteria Metric (top in black and white; bottom is color coded to indicate Category rankings).

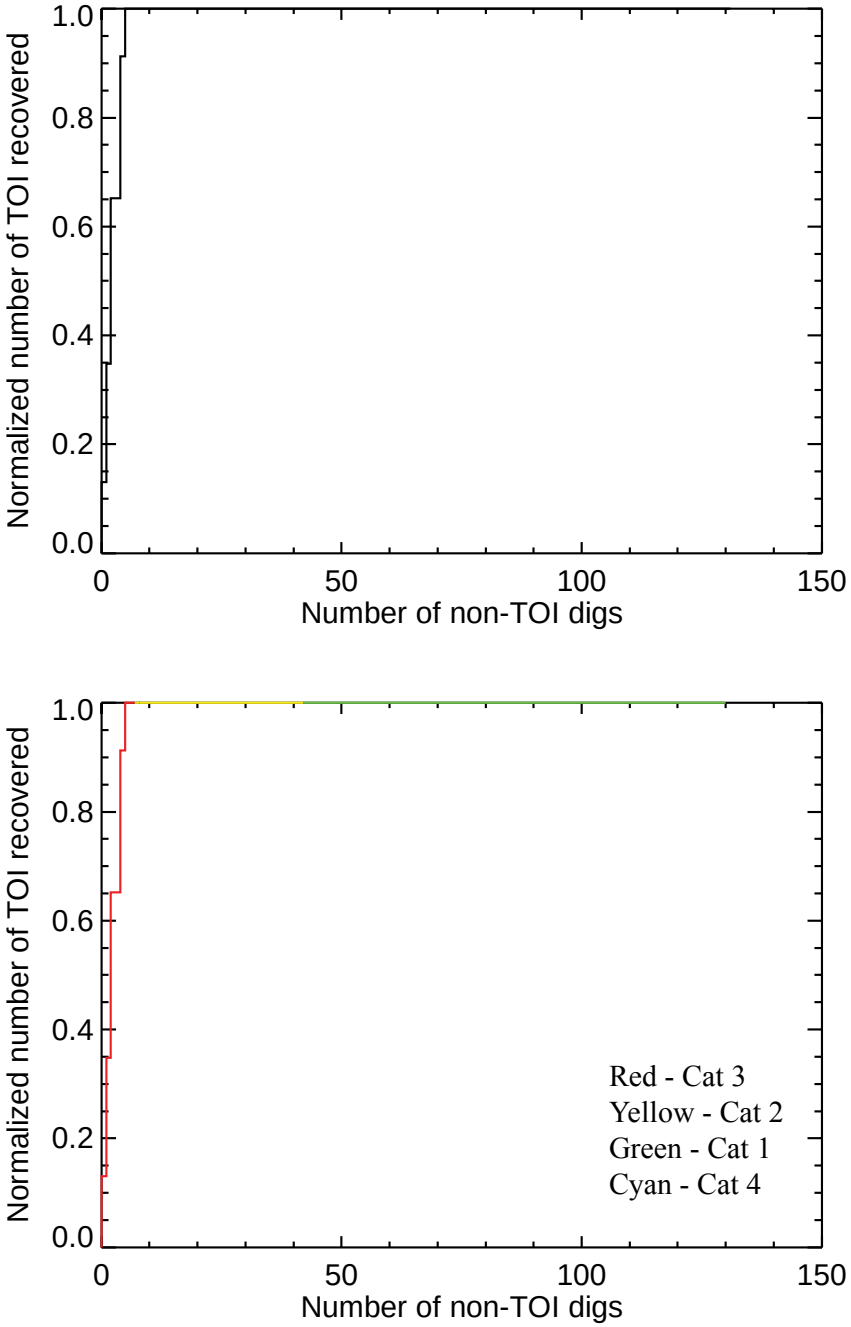


Figure 6-43. Metal Mapper ROC curve for the 2 Criteria Metric (top in black and white; bottom is color coded to indicate Category rankings).

6.4 OVERLAPPING SIGNATURES

Overlapping signatures was a serious problem at Camp SLO. If not recognized and handled appropriately, overlapping signatures often result in inaccurate target attributes and potential false negatives. Recognizing and handling overlapping signatures was not a simple task.

There were some straightforward measures that were used. The data-to-model mismatch was used, for example, to identify poor fits – which may result from overlapping signatures. If the mismatch is large, the fitted parameters were not trusted. The spatial footprint and general shape of the anomalies were also be used to identify overlapping signatures. These approaches required an aggressive quality-control process that evaluated each and every anomaly. The features that we tracked include sample distance, time sampling, heading, elevation, spatial footprint, lane separation, orientation, signal decay, late to early decay rate, change of beta values with fit depth, amplitude differences for overlapping locations and fit error for cases where we limit the input data. We generated these statistics and created QC graphical plots (see Figure 6-44) for all EM61 Master List targets. We saved the statistics for each anomaly analyzed. Flags were used to indicate which criteria failed and thus which anomalies require manual review. If the target data do not pass our QC tests, the data cannot be used to justify a do-not-dig decision. This was the first measure against overlapping signatures.

Another approach to identifying overlapping signatures is to fit the anomaly data using multiple source models and comparing the results. A multi-source situation may be present if a given anomaly is much better fit using two or more dipoles instead of one. An advantage of this solver is that it can be batch run and does not necessarily require another exhaustive look by the analyst. We used a multi-source solver as a second measure against overlapping signatures for the EM61 and magnetic data.

The fixed, multi-coil construction of the TEMTADS and Metal Mapper allow for a more aggressive approach. For these sensors, SAIC carried out multi-target inversions using a new algorithm [12] which permits several sources to be characterized simultaneously. The algorithm proceeds by iteratively updating a large collection of hypothesized sources, assumed to be far greater in number than the actual sources, and forming linear combinations of their associated signals, calculated through the forward model, to approximate the observed signal y. With each iteration we solve $y=Ax$, in which each column of matrix A contains the entire expected signal from a single hypothesized source, arranged as a column vector and normalized to unit magnitude. The solver delivers the solution vector x, containing mostly zeros, which are the weights applied to those signals in order to best-match the observed data y. The algorithm operates by refining the collection of sources according to these weight assignments x, and using knowledge of relationships between source parameters and associated signals to iteratively home-in on the true distribution of sources.

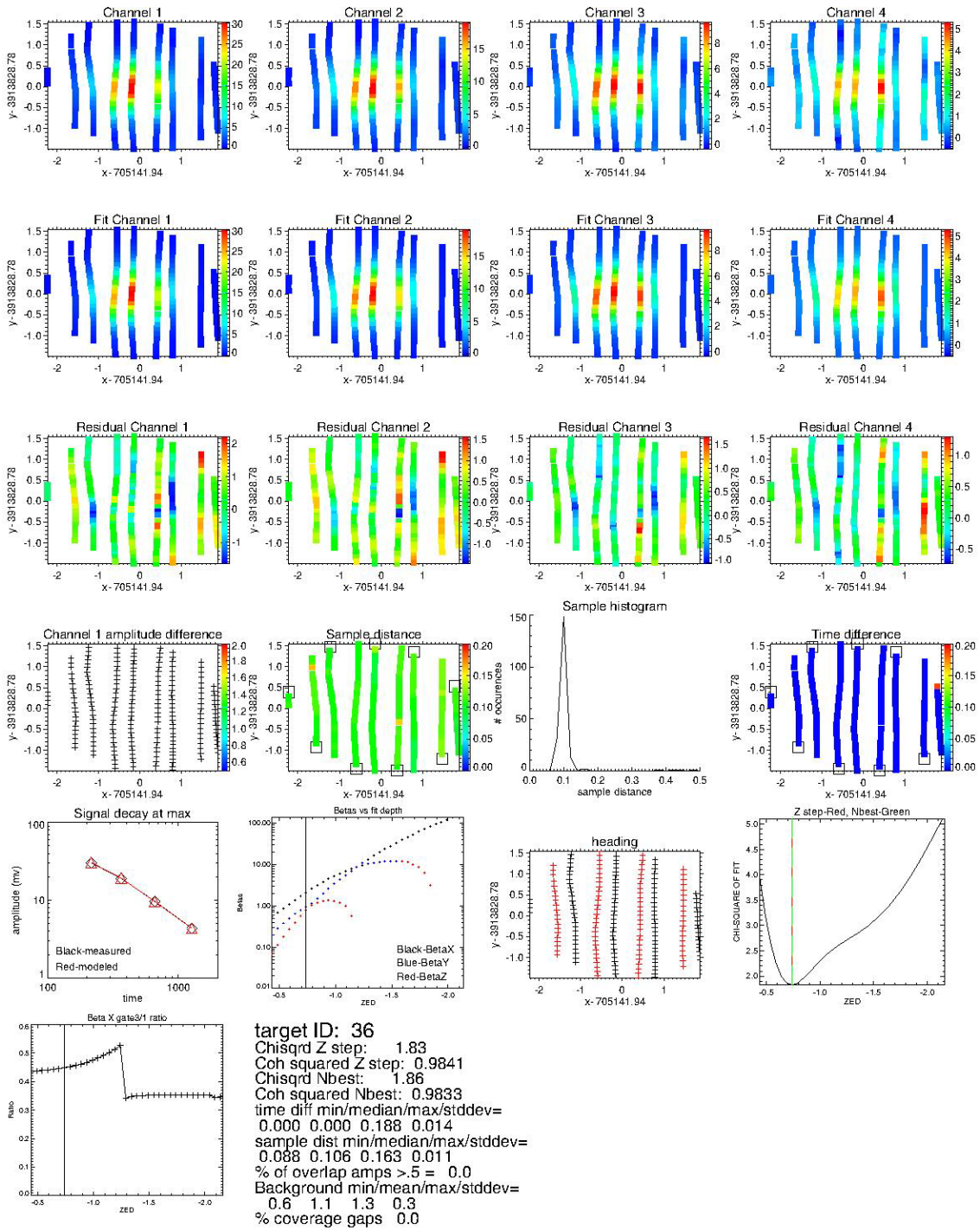


Figure 6-44. Quality control plot for Master ID 36 – a 60mm buried 25cm below grade. This anomaly has a fairly clean CHI-Squared surface (bottom right plot) and displays axial symmetry at the minimum error.

In this implementation, spatial locations of sources were restricted to a lattice of discrete nodes arranged in 24 horizontal planes, each having a regular square tiling in the plane. To ensure that differences in signal between adjacent nodes are uniform, the lattice becomes coarser toward the bottom where signals change less rapidly with target location. Example results of this approach are presented in Figure 6-45 and Figure 6-46.

We batch fit all TEMTADS and Metal Mapper data using the new multi-target solver. We used the results to augment our standard inversion approach by incorporating the inverted multi-source polarizabilities when appropriate.

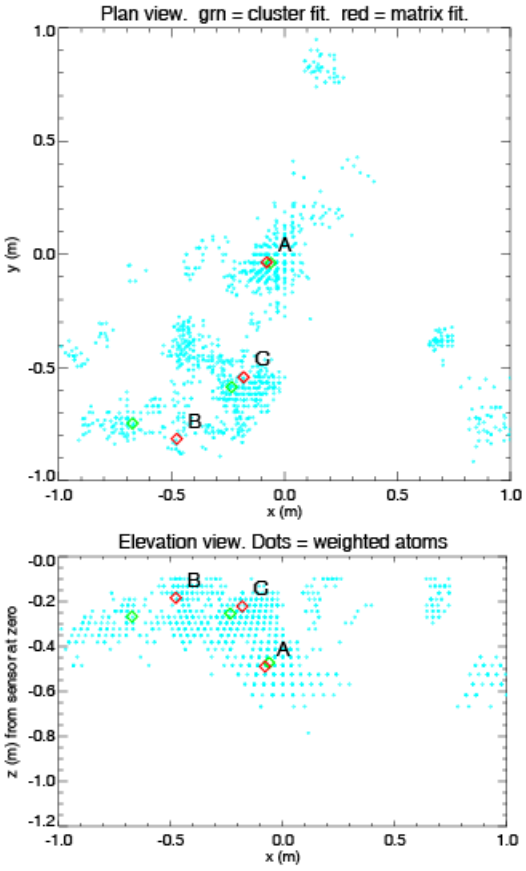


Figure 6-45. Results of SAICs multi-target solver are presented in plan and elevation view where the symbols represent weights associated with assumed dipoles that occupy a subsurface lattice structure. Data from APG blind grid (H5).

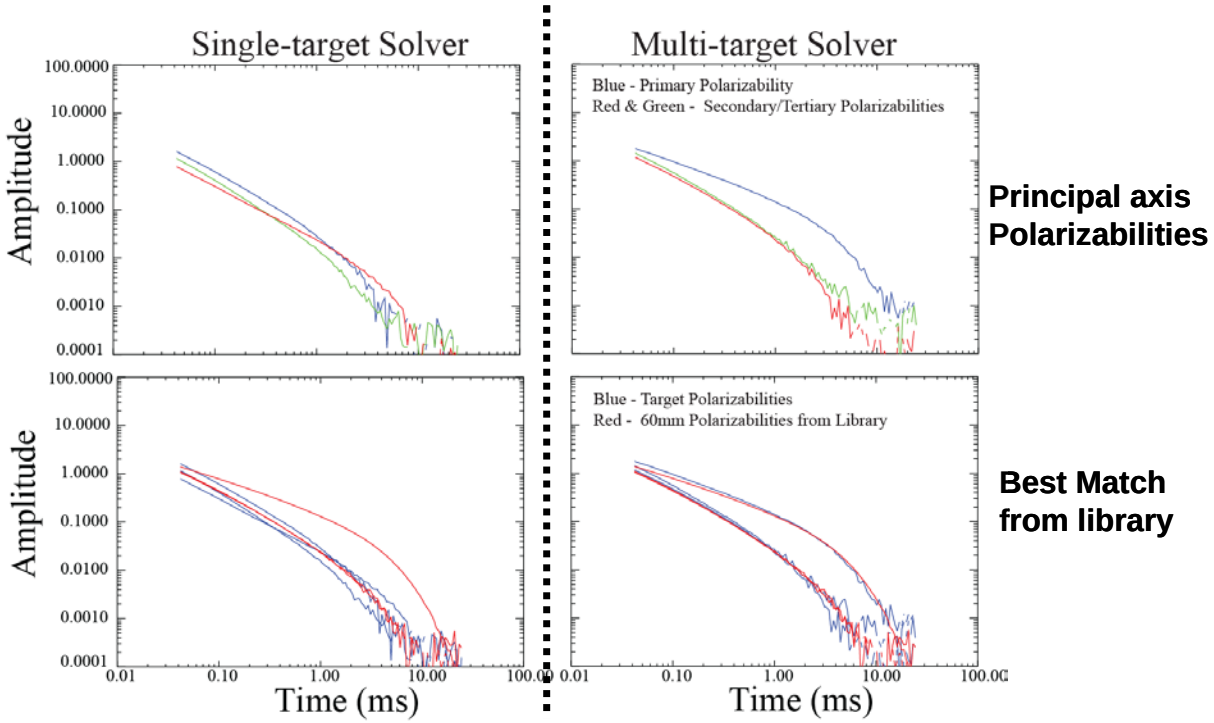


Figure 6-46. Comparison of principal axis polarizabilities for APG H5, which contains a buried 60mm mortar. The single target solution shown on the left does not accurately indicate a buried 60mm mortar. The multi-target solution, shown on the right, does recover polarizations that match 60mm mortars.

7 PERFORMANCE ASSESSMENT

All eight submitted dig lists were scored against the emplaced and recovered targets by the IDA. An example ROC curve is shown in Figure 7-1, with the areas of interest for the analysis indicated.

The scoring software plotted the Percent of Munitions Dug versus the Number of Unnecessary Digs for all possible dig thresholds and drew vertical grey bars around each point on the ROC curve to denote the 95% confidence interval around the point's Percent of Munitions Dug value. From a declared-category perspective, the Category 4 targets are plotted first, followed by categories 3 (red), 2 (yellow), and finally 1 (green). The colored dots on the ROC curves indicate the operating point for a Pd=0.95 (pink), the demonstrator's threshold point (dark blue) and the "best case scenario" dig threshold which has the lowest number of FP for Pd=1.0 (light blue). The scoring software colored in black the extra point on the ROC curve corresponding to the "Dig Everything" situation, in which ALL locations in the Test Set are declared "Dig"; this point is always at the upper, right end of the ROC curve. Conversely, the software also colored in black the extra point corresponding to the "Leave Everything in the Ground" situation, in which ALL locations in the Test Set are declared "Do not dig", including the Cannot Analyze locations. This point is always at the lower, left end of the ROC curve.

In a real-world situation, all locations in the Training Set must be dug. Training set locations that were truly of the munition type of interest cause a constant shift increase in the Percent of Munitions Dug values plotted along the Y axis. Similarly, training set locations that were truly Clutter cause a constant shift increase in the Number of Unnecessary Digs values plotted along the X axis. Including the training set locations allows for an apples-to-apples comparison between ranked dig lists created from training sets of different sizes, as all ROC curves based on the same data collection instrument will have their upper, right "Dig Everything" point in the same location, regardless of the size of the training set used.

All locations in the Test Set that the demonstrators declared to be "Category 4 = Cannot Analyze" must also be dug. The Cannot Analyze locations in the Test Set are represented by the vertical and horizontal offsets between the lower, left end of the ROC curve and the "Leave Everything in the Ground" black dot:

ROC curves were generated for each anomaly list submitted. Each data set will be discussed individually in the sections that follow. As part of the discussion a number of figures are presented comparing the inverted parameters to ground truth information for all excavated anomalies. Due to the large number of anomalies, we have segmented the plots based on our classification categories. This also serves to illustrate differences between the categories. In all figures, UXO/TOI are plotted in red, while clutter are plotted in black. In general the agreement with ground truth improves with increasing category number. This is presumed to be due to a large number of small, low SNR frag items in Category 1, which result in greater uncertainty in both the measured and fitted values. Also, UXO objects being compact in size and shape tend to fit our point dipole model better than some clutter objects which can be quite irregular in shape.

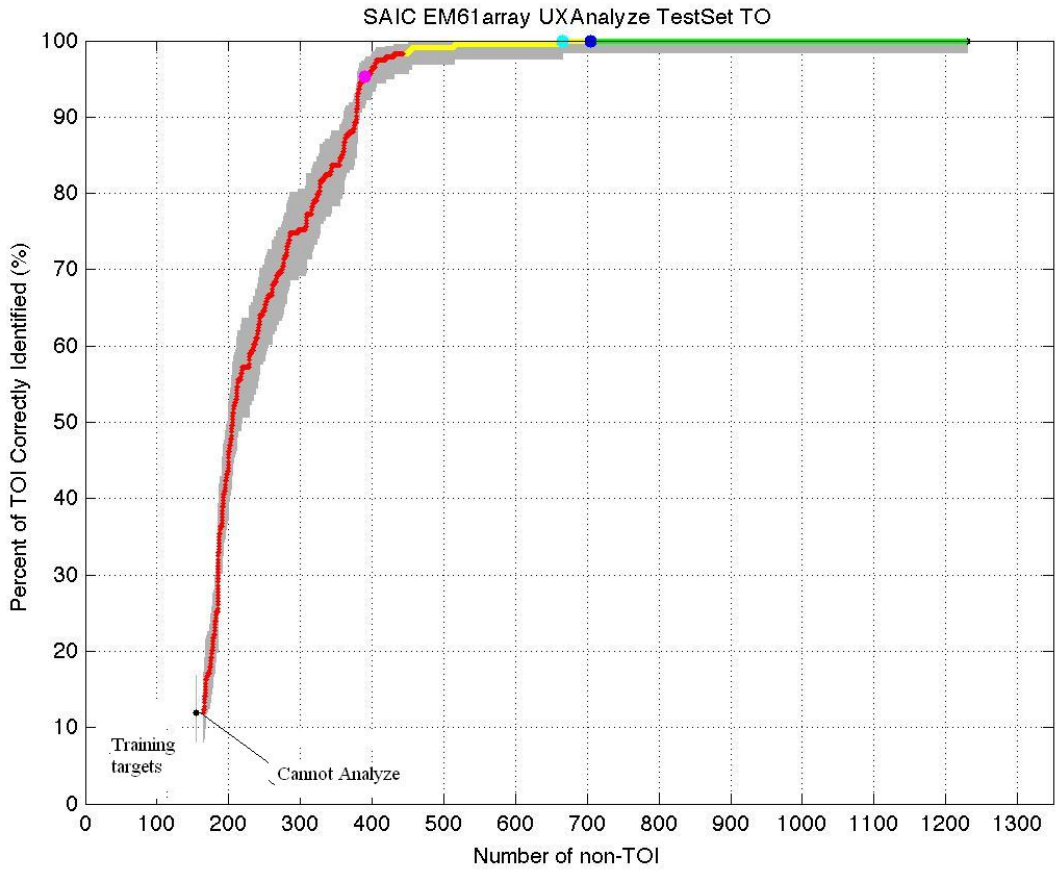


Figure 7-1. Example ROC curve with areas of interest indicated. The colored dots on the ROC curves indicate the operating point for a Pd=0.95 (pink), the demonstrator's threshold point (dark blue) and the "best case scenario" dig threshold which has the lowest number of FP for Pd=1.0 (light blue)..

7.1 SLOPE CORRECTED EM61 MK2 CART

Data from the SCORR EM61 MK2 cart is shown in Figure 7-2. The circles identify anomalies selected for Test Set by the ESTCP Program Office while the squares represent the anomalies selected for training the classifiers.

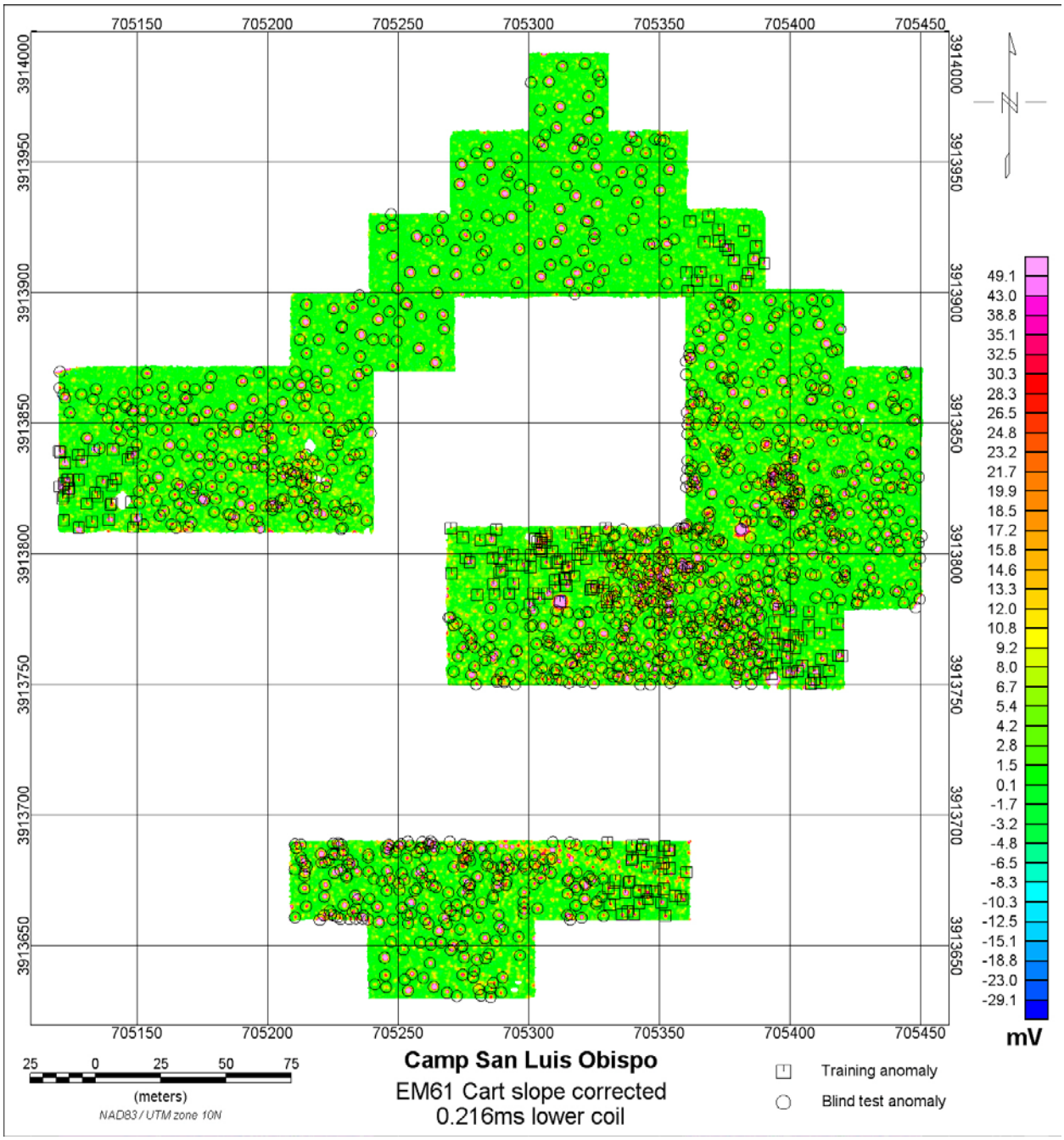


Figure 7-2. Slope corrected EM61 MK2 Cart data.

7.1.1 Performance Scores from IDA

Scoring performances for the SCORR EM61 MK2 cart analysis are reported in Table 7-1 and shown graphically in Figure 7-3. A ROC chart is shown in Figure 7-4, where we plot the Percent of Munitions Dug versus the Number of Unnecessary Digs.

Using the thresholds adopted for this analysis, there was one false negative. Anomaly #1444, which was a seeded 60mm body, was classified as high confidence clutter (Figure 7-5).

Table 7-1 Test Set Summary: Slope corrected EM61 MK2 Cart

Category	Cultural	Munition Debris	No Contact	Soil	UXO
1	21	586	6	0	1
2	4	247	12	0	11
3	10	180	7	0	200
4	0	1	0	0	0
TOTAL	**35**	**1014**	**25**	**0**	**212**

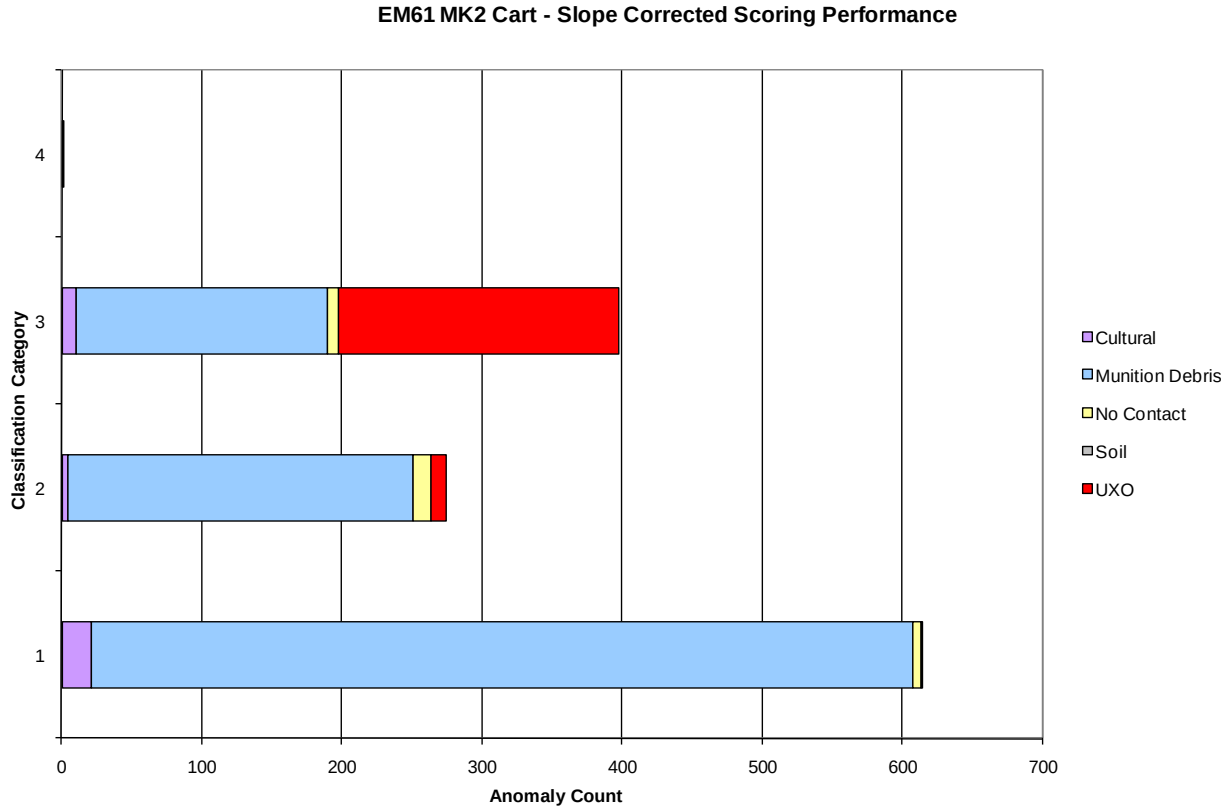


Figure 7-3. SCORR EM61 MK2 Cart performance as a function of classification category.

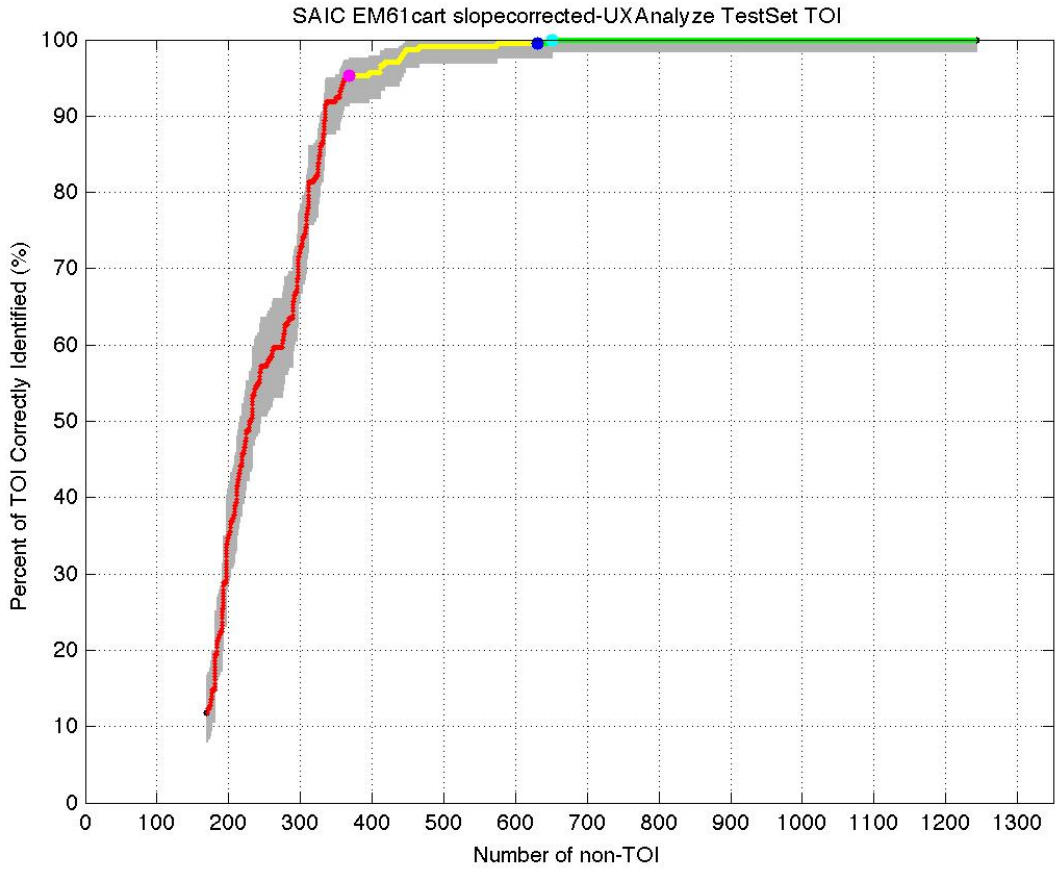


Figure 7-4. SCORR EM61 MK2 Cart ROC chart.

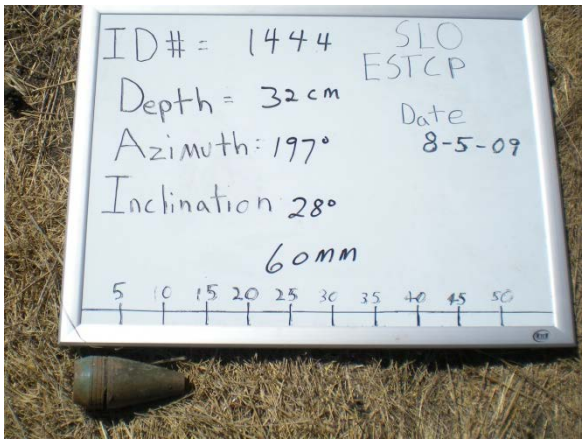


Figure 7-5. Photograph of target 1444 which is a 60mm body that was classified as high confidence clutter.

7.1.2 Characterization Plots

Figure 7-6 shows the difference between the fitted and measured XY locations for all category 1, 2 and 3 targets from the Test Set. The mean error for all TOI with an isolated or slightly overlapping signal was 0.22m with a standard deviation of .131m. If non-TOI are added to the population the mean error increases to 0.28m with a standard deviation of 0.21m.

Figure 7-7 shows the difference between the fitted and true depth for all category 1, 2 and 3 targets from the Test Set. The mean error for all TOI with an isolated or slightly overlapping signal was -0.13m with a standard deviation of .10m. If non-TOI are added to the population the mean error increases to -0.24m with a standard deviation of 0.16m.

Figure 7-8 and Table 7-2 show the inverted polarizabilities for all category 1, 2 and 3 targets from the Test Set. In general, the calculated size (sum of betas) of the main TOI shows much smaller relative deviations than the individual betas. The relative deviations are also typically smaller for the larger TOI (4.2in mortar) which also have less UXO variability.

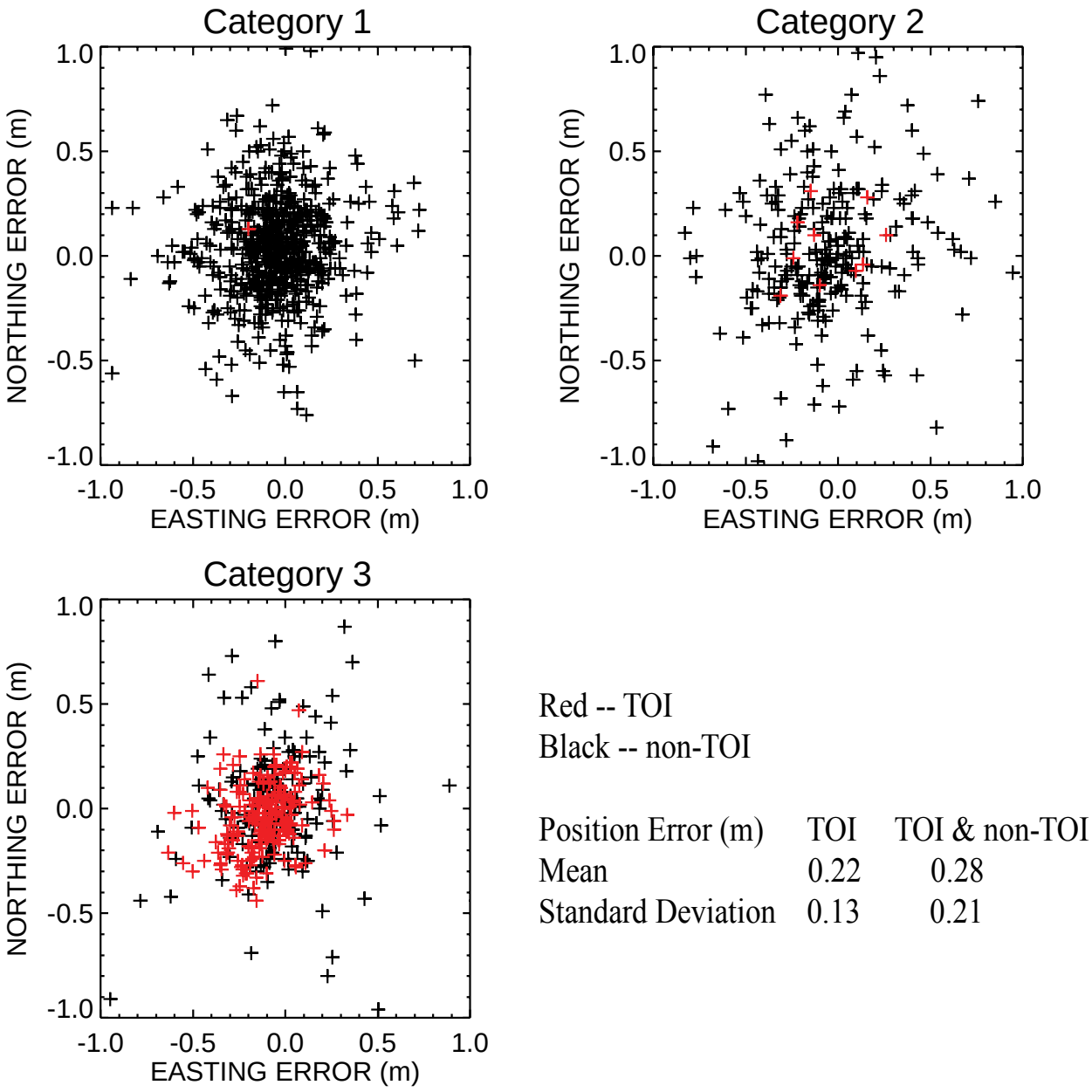


Figure 7-6. Differences between fitted and measured XY locations; SCORR EM61 MK2 cart

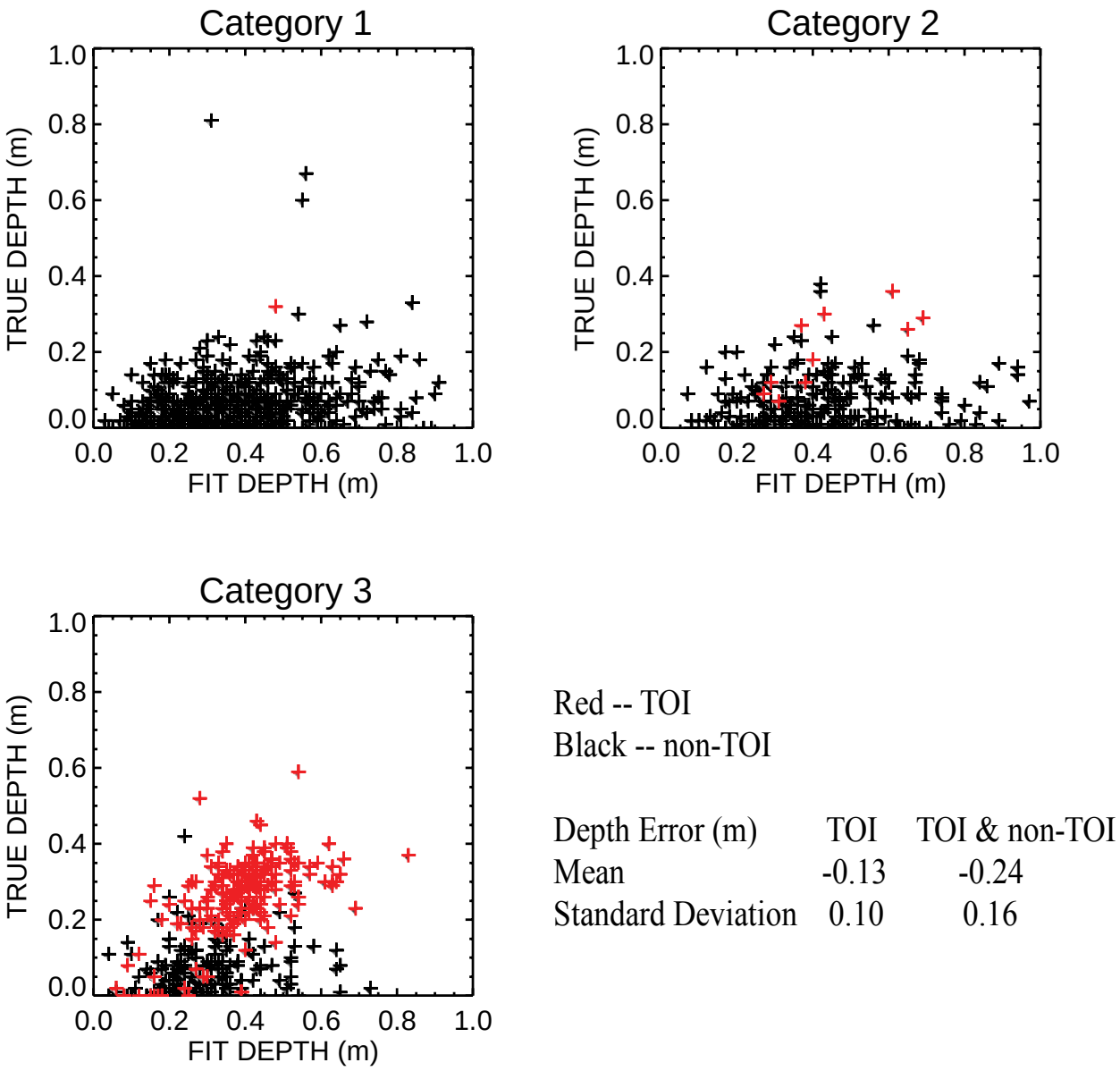


Figure 7-7. Fitted versus measured depth of burial; SCORR EM61 MK2 cart

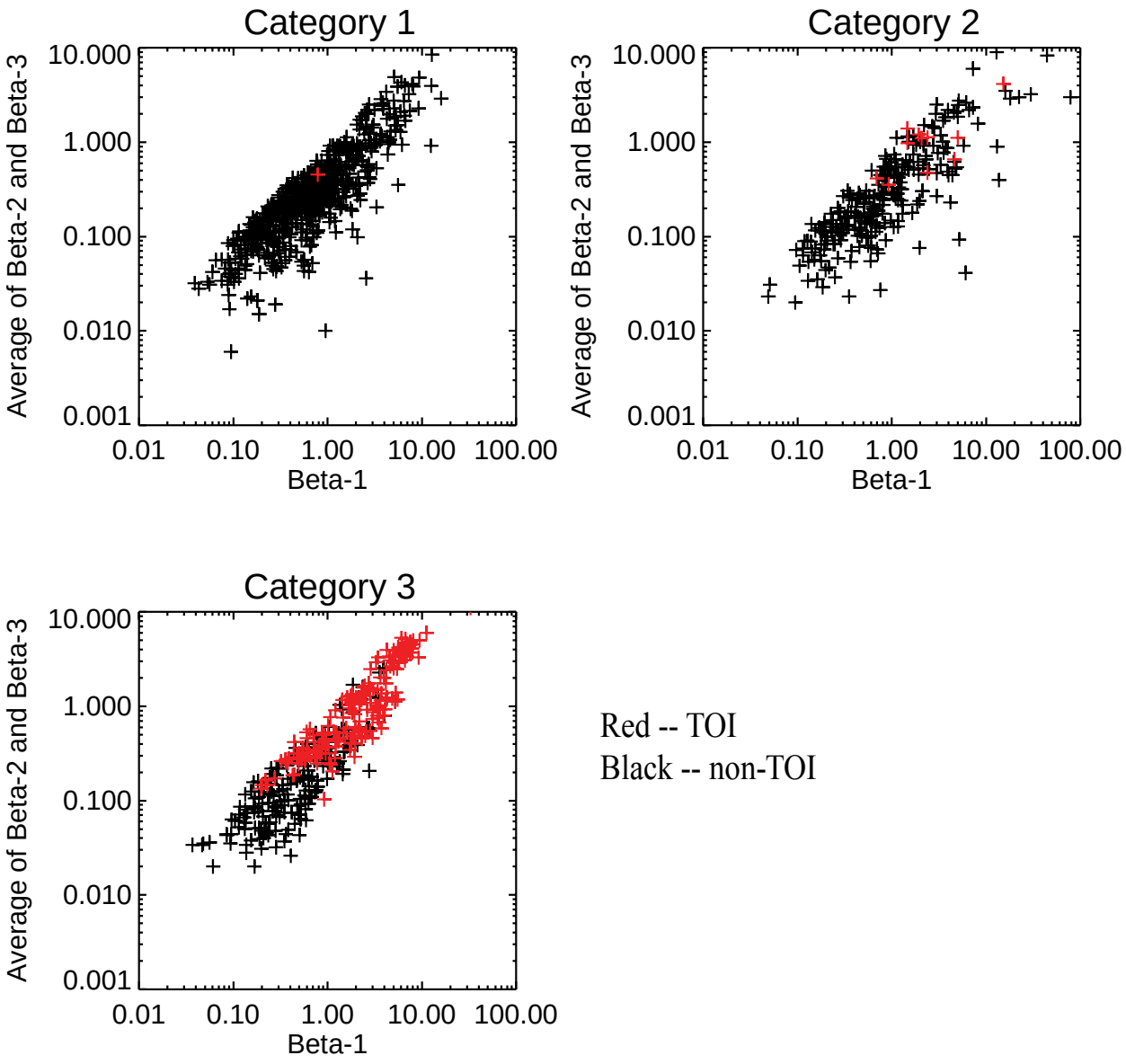


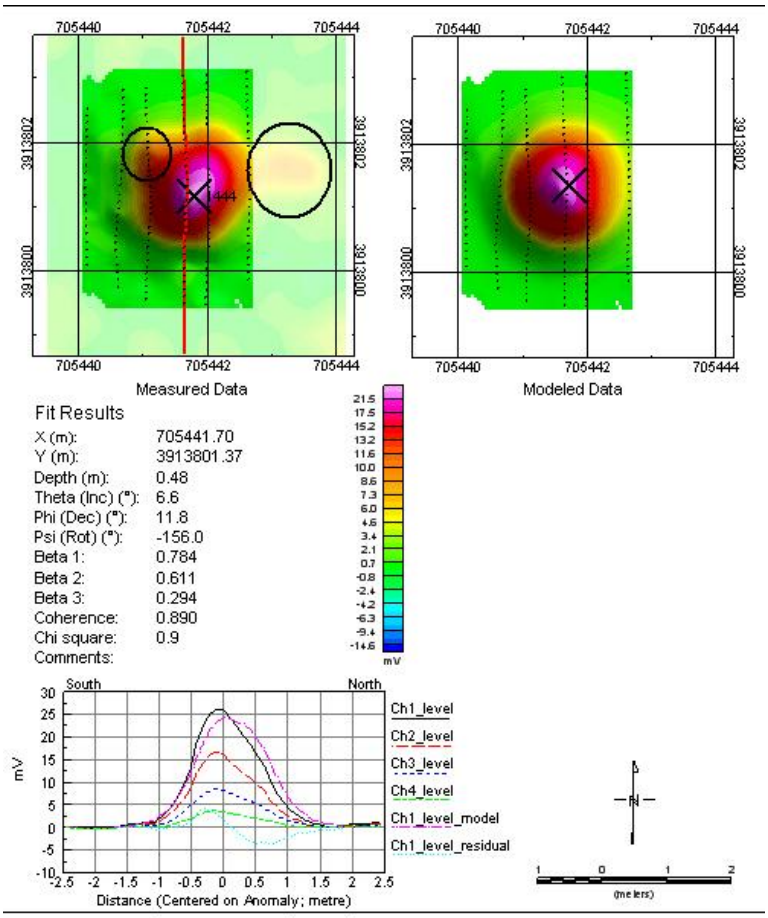
Figure 7-8. Beta 1 versus the average of Beta 2 and Beta 3; SCORR EM61 MK2 cart

Table 7-2 Statistics of betas for the four main TOI, SCORR EM61 MK2 cart

Type	# of samples	Size		Beta 1		Beta 2		Beta 3	
		Mean	Std Dev	Mean	Std Dev	Mean	Std Dev	Mean	Std Dev
60mm	74	0.057	0.011	1.055	0.863	0.412	0.250	0.174	0.133
81mm	56	0.084	0.012	2.927	1.644	1.329	0.808	0.705	0.340
2.36in	21	0.076	0.009	2.515	0.826	0.663	0.311	0.384	0.176
4.2in	51	0.115	0.009	6.074	1.349	3.740	0.772	2.598	0.712

7.1.3 Failure Analyses

Anomaly #1444 was a seeded 60mm body but was classified as high confidence clutter (Category 1). It had a decision metric of 0.066 which was slightly below our category 1-2 threshold of 0.1. A reexamination of the anomaly data revealed that the size parameter indicated a high confidence TOI but the decay parameter indicated a high confidence non-TOI. The gate 4 to gate 1 ratio was 0.138 which was smaller than any of the 60mm in our training set which had values ranging from 0.162 to 0.230. Figure 7-9 shows the measured and modeled data for anomaly 1444. On careful examination, it appears that there may be some small sources near the anomaly (indicated by the circles) that affected the data sufficiently enough to cause the failure. The maximum amplitude of this anomaly was 26mV for the first time gate. A small fast decaying object with an amplitude of about 5mV would be enough to reduce that calculated decay ratio to the value observed.



SLO - EM61 cart - Anomaly 1444

Figure 7-9. Anomaly plot showing measured data, inverted features and forward model for anomaly 1444 which was ranked just below the Category 1-2 boundary.

7.2 NON SLOPE CORRECTED EM61 MK2 CART

Data from the non SCORR EM61 MK2 cart is shown in Figure 7-10. The circles identify anomalies selected for the Test Set by the ESTCP Program Office while the squares represent the anomalies selected for training the classifiers.

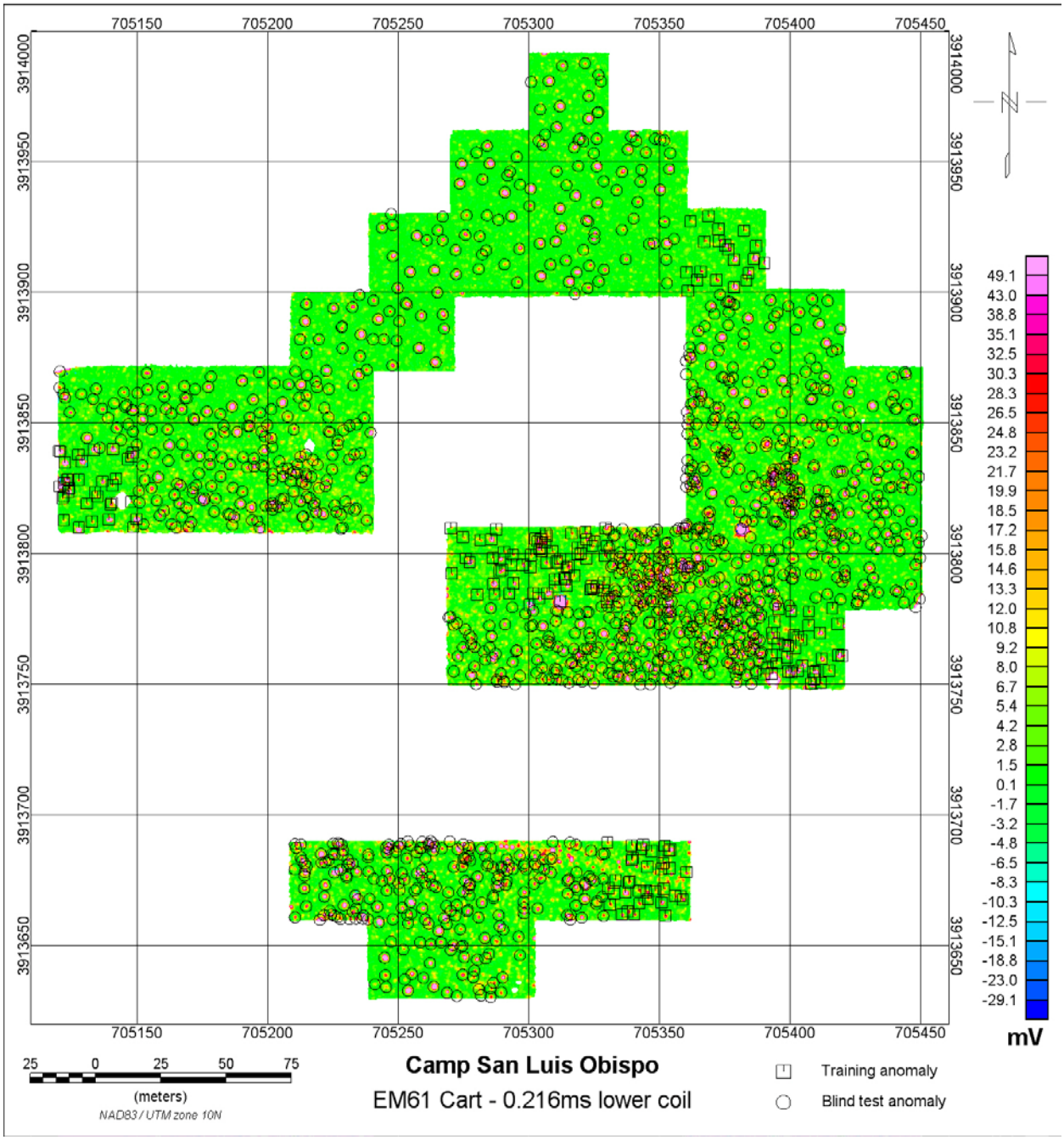


Figure 7-10. Non SCORR EM61 MK2 Cart data.

7.2.1 Performance Scores from IDA

Scoring performances for the non SCORR EM61 MK2 cart analysis are reported in Table 7-3 and shown graphically in Figure 7-11. A ROC chart is shown in Figure 7-12, where we plot the Percent of Munitions Dug versus the Number of Unnecessary Digs.

Using the thresholds adopted for this analysis, there was one false negative. As with the SCORR data anomaly #1444 was classified as high confidence clutter.

Table 7-3 Test Set Summary: Non SCORR EM61 MK2 Cart

Category	Cultural	Munition Debris	No Contact	Soil	UXO
1	21	590	4	0	1
2	4	251	14	0	15
3	10	171	7	0	196
4	0	2	0	0	0
TOTAL	**35**	**1014**	**25**	**0**	**212**

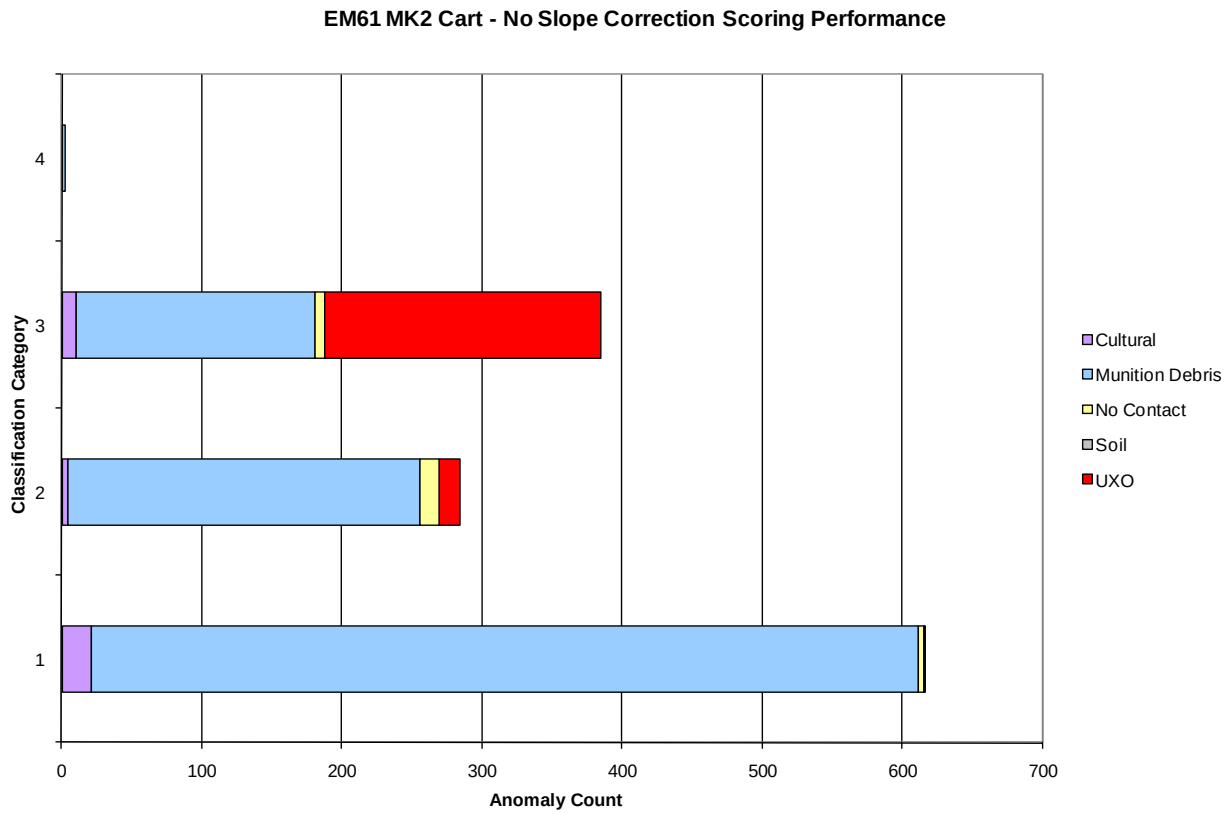


Figure 7-11. Non SCORR EM61 MK2 Cart performance as a function of classification category.

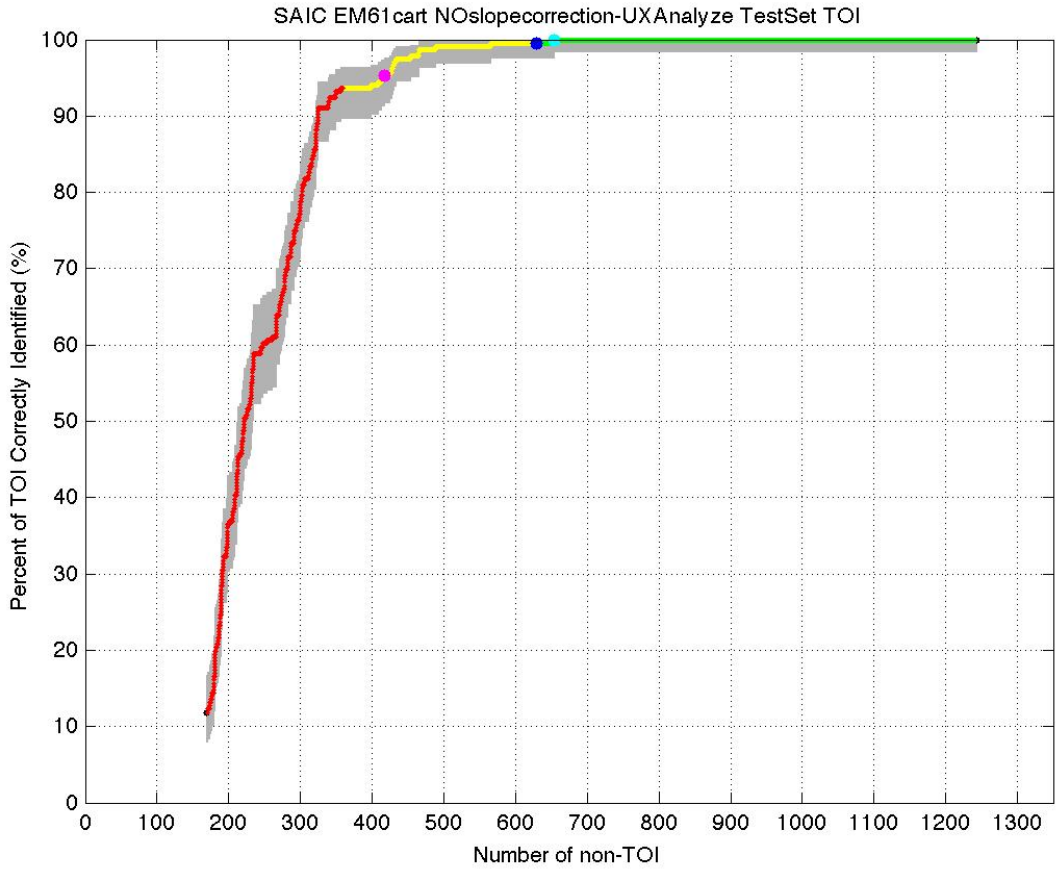


Figure 7-12. Non SCORR EM61 MK2 Cart ROC chart.

7.2.2 Characterization Plots

Figure 7-13 shows the difference between the fitted and measured XY locations for all category 1, 2 and 3 targets from the Test Set. The mean error for all TOI with an isolated or slightly overlapping signal was 0.32m with a standard deviation of .17m. If non-TOI are added to the population the mean error increases to 0.35m with a standard deviation of 0.22m.

Figure 7-14 shows the difference between the fitted and true depth for all category 1, 2 and 3 targets from the Test Set. The mean error for all TOI with an isolated or slightly overlapping signal was -0.13m with a standard deviation of .11m. If non-TOI are added to the population the mean error increases to -0.24m with a standard deviation of 0.17m.

Figure 7-15 and Table 7-4 show the inverted polarizabilities for all category 1, 2 and 3 targets from the Test Set. In general, the calculated size (sum of betas) of the main TOI shows much

smaller relative deviations than the individual betas. The relative deviations are also typically smaller for the larger TOI (4.2in mortar).

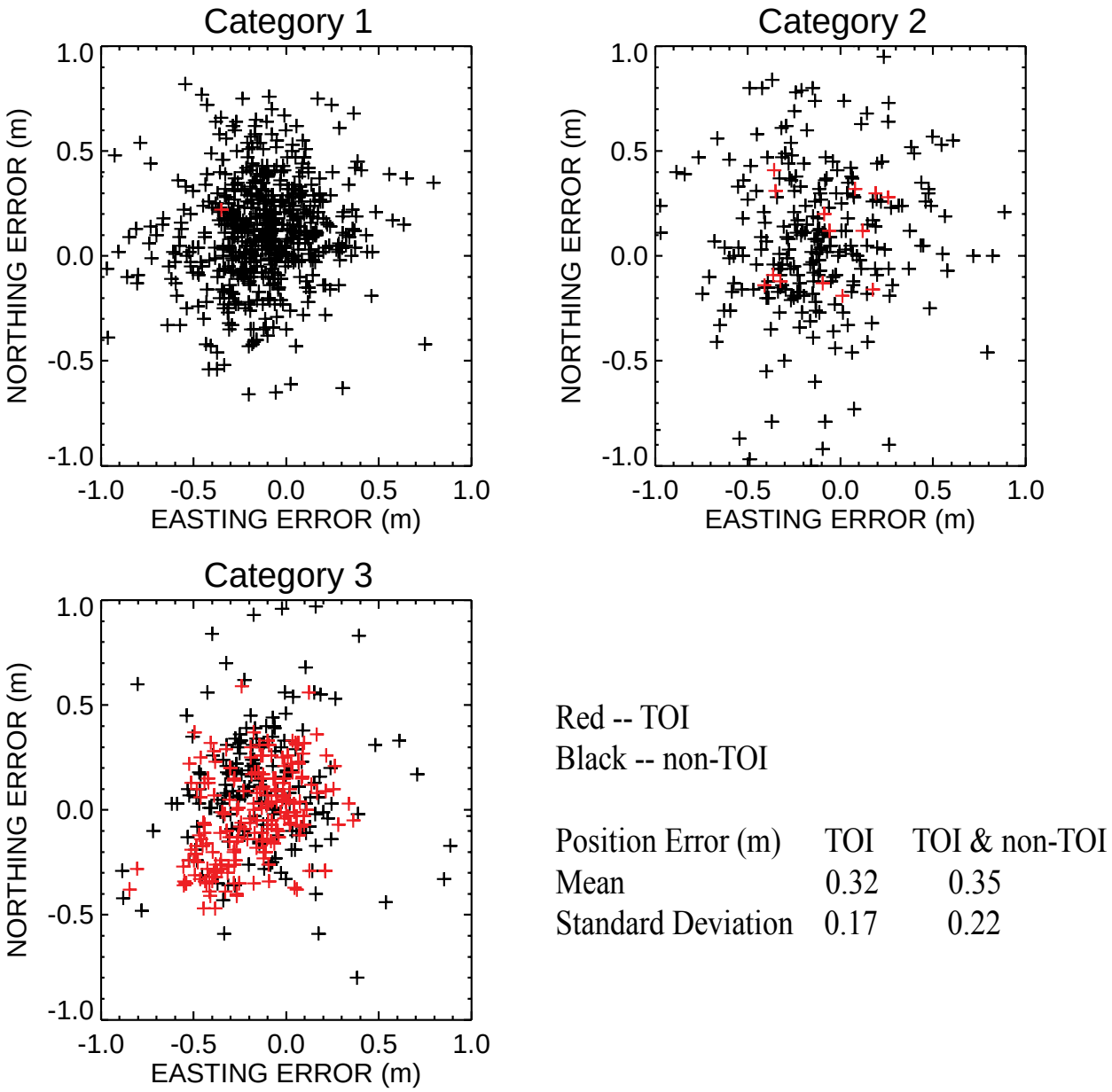


Figure 7-13. Differences between fitted and measured XY locations; non SCORR EM61 MK2 cart

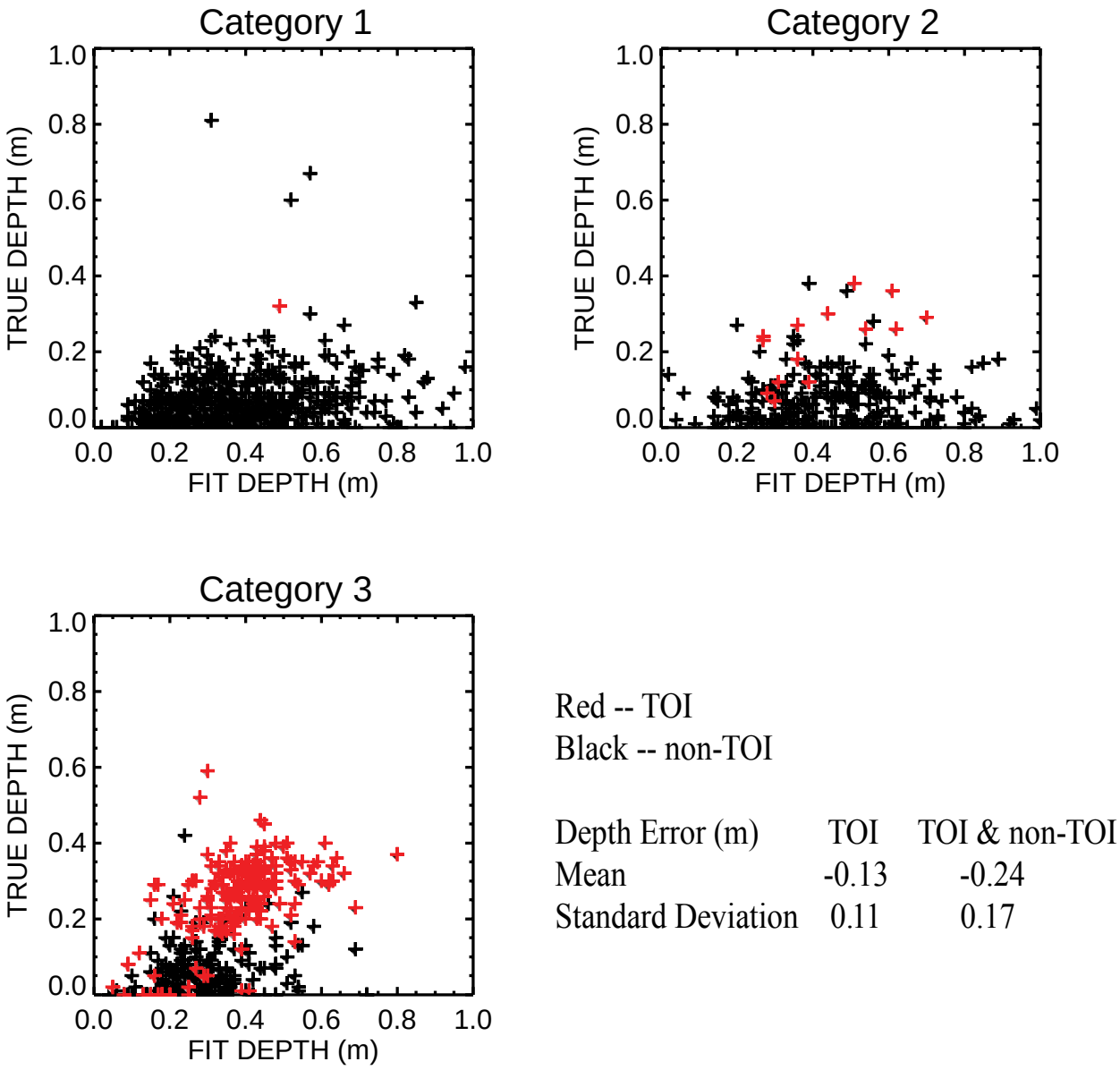


Figure 7-14. Fitted versus measured depth of burial; non SCORR EM61 MK2 cart

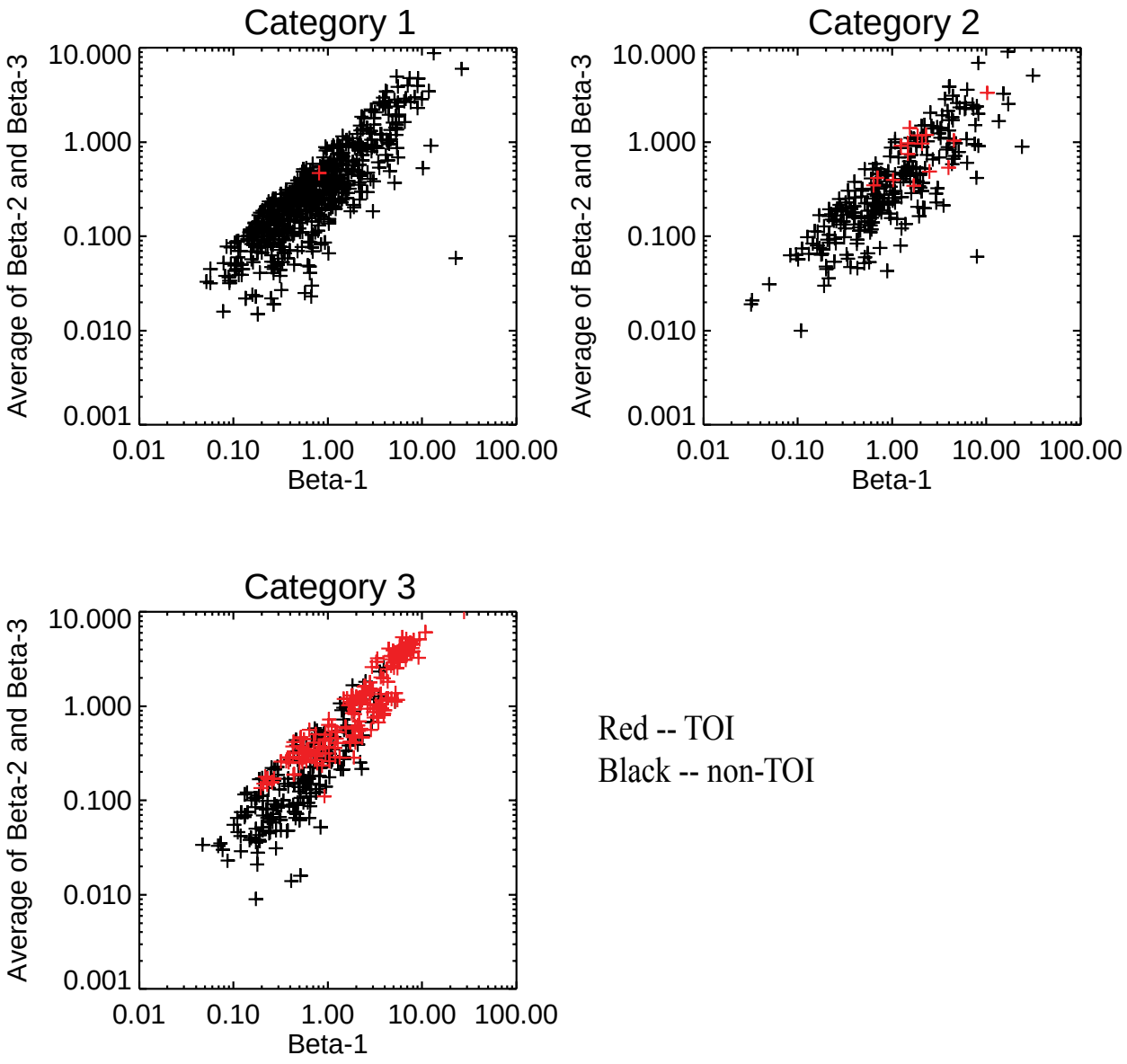


Figure 7-15. Beta 1 versus the average of Beta 2 and Beta 3; non SCORR EM61 MK2 cart

Table 7-4 Statistics of Betas for the four main TOI, non SCORR EM61 MK2 cart

Type	# of samples	Size		Beta 1		Beta 2		Beta 3	
		Mean	Std Dev	Mean	Std Dev	Mean	Std Dev	Mean	Std Dev
60mm	73	0.056	0.011	0.951	0.718	0.407	0.249	0.181	0.135
81mm	56	0.083	0.013	2.896	1.642	1.333	0.825	0.711	0.347
2.36in	21	0.075	0.010	2.528	0.929	0.671	0.330	0.396	0.185
4.2in	51	0.115	0.009	6.111	1.341	3.764	0.775	2.615	0.717

7.2.3 Failure Analyses

Anomaly #1444 was a seeded 60mm body but was classified as high confidence clutter (Category 1). This was the same failure as seen in the SCORR EM61 MK2 data. The cause of the failure is the same as discussed in the section describing the SCORR data.

7.3 EM61 ARRAY

Data from the EM61 Array is shown in Figure 7-16. The circles identify anomalies selected for the Test Set by the ESTCP Program Office while the squares represent the anomalies selected for training the classifiers.

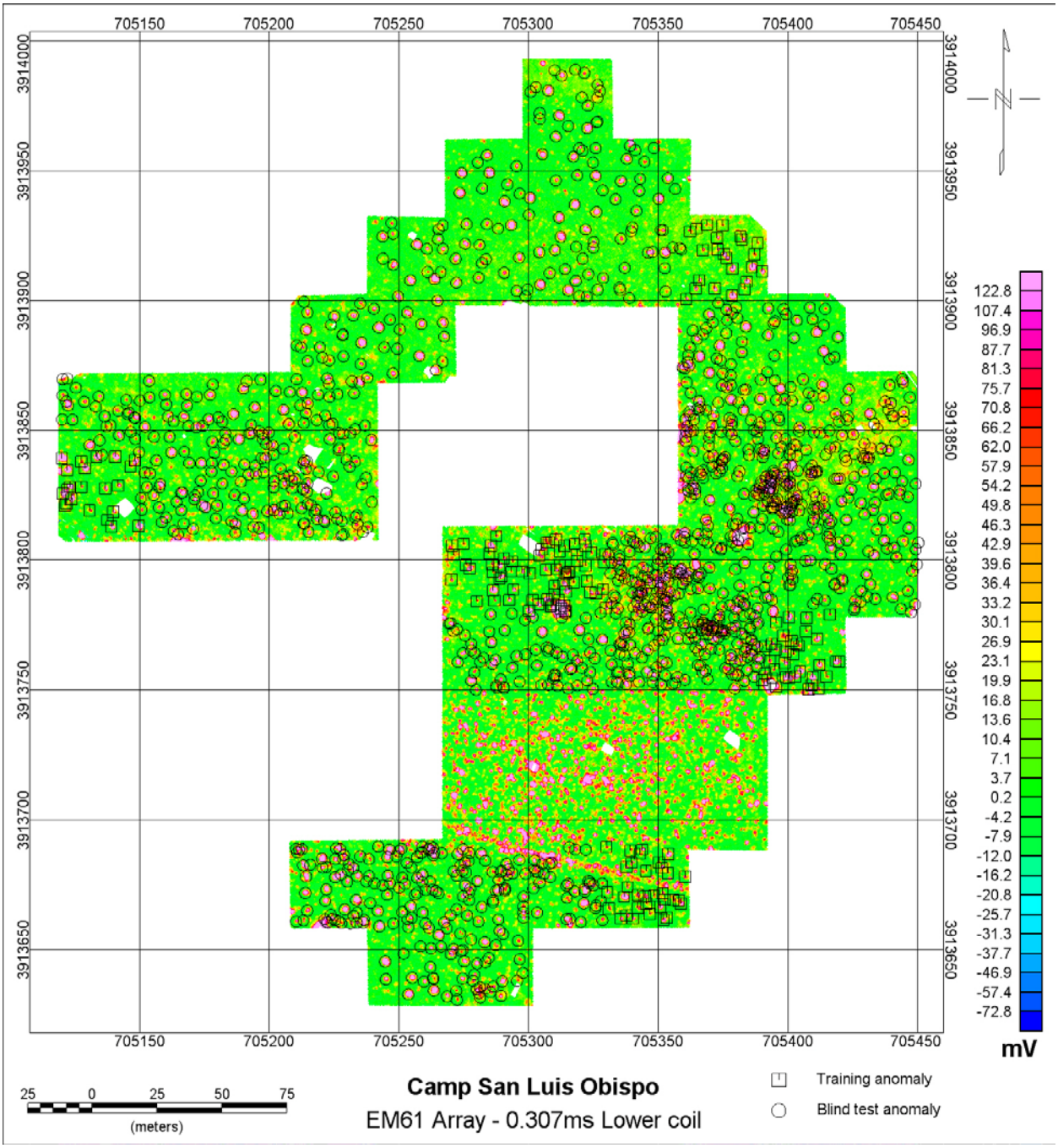


Figure 7-16. EM61 Array data.

7.3.1 Performance Scores from IDA

Scoring performances for EM61 Array analysis are reported in Table 7-5 and shown graphically in Figure 7-17. A ROC chart is shown in Figure 7-18, where we plot the Percent of Munitions Dug versus the Number of Unnecessary Digs.

Using the thresholds adopted for this analysis, there were no false negatives.

Table 7-5 Test Set Summary: EM61 Array

Category	Cultural	Munition Debris	No Contact	Soil	UXO
1	17	502	6	0	0
2	8	234	18	1	4
3	11	254	14	0	202
4	1	8	2	0	0
TOTAL	**37**	**998**	**40**	**1**	**206**

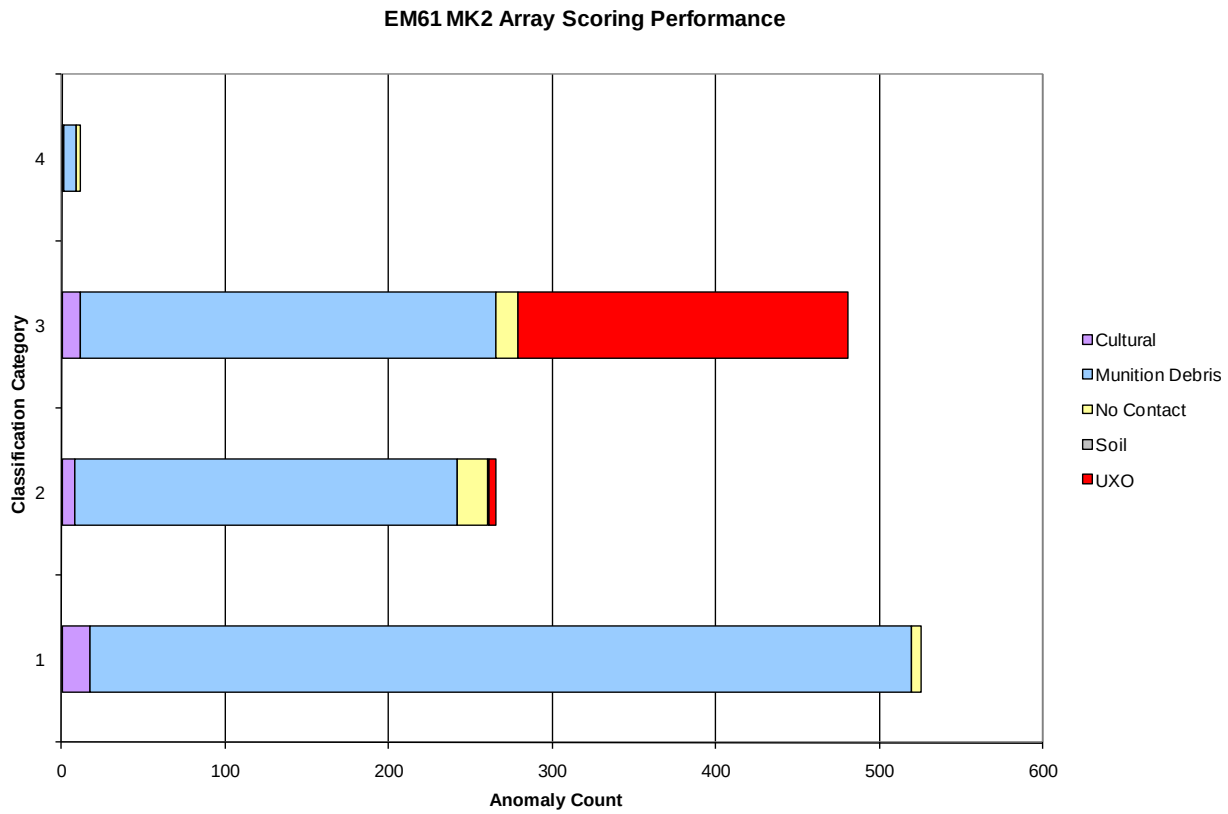


Figure 7-17. EM61 Array performance as a function of classification category.

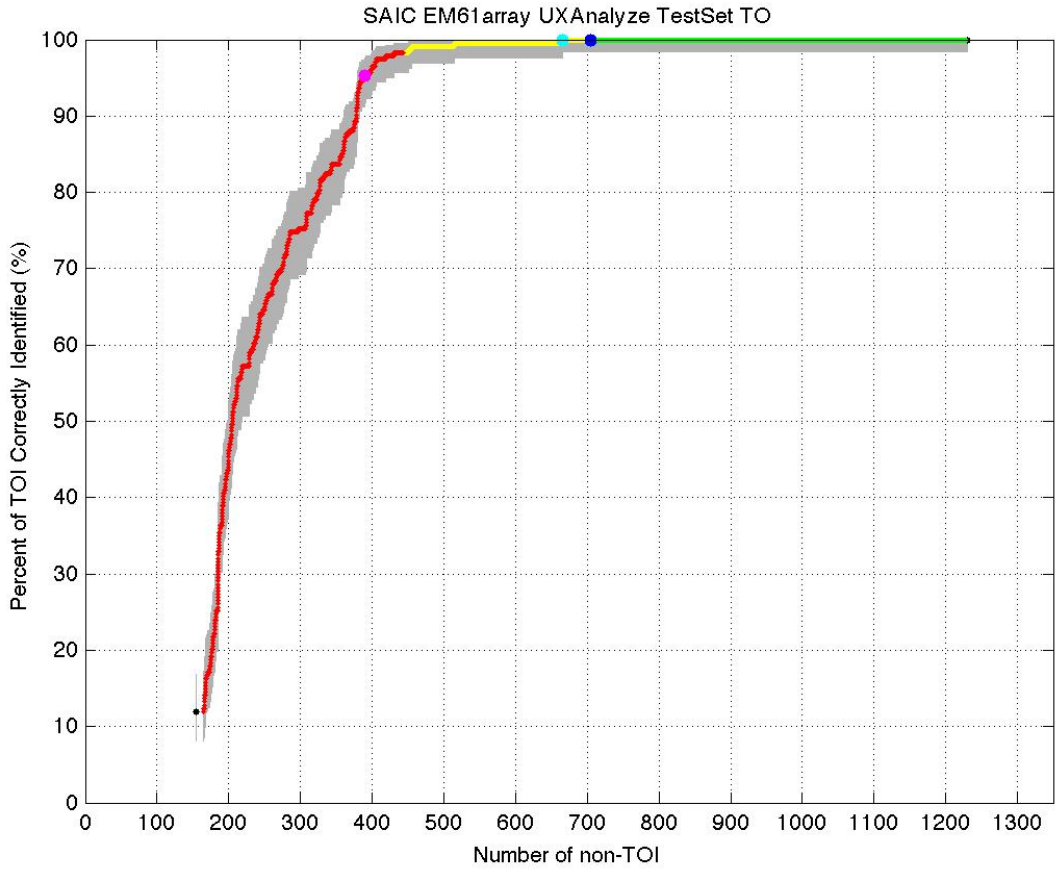


Figure 7-18. EM61 Array ROC chart.

7.3.2 Characterization Plots

Figure 7-19 shows the difference between the fitted and measured XY locations for all category 1, 2 and 3 targets from the Test Set. The mean error for all TOI with an isolated or slightly overlapping signal was 0.24m with a standard deviation of 0.14m. If non-TOI are added to the population the mean error increases to 0.28m with a standard deviation of 0.23m.

Figure 7-20 shows the difference between the fitted and true depth for all category 1, 2 and 3 targets from the Test Set. The mean error for all TOI with an isolated or slightly overlapping signal was -0.19m with a standard deviation of 0.12m. If non-TOI are added to the population the mean error increases to -0.22m with a standard deviation of 0.22m.

Figure 7-21 and Table 7-6 show the inverted polarizabilities for all category 1, 2 and 3 targets from the Test Set. In general, the calculated size (sum of betas) of the main TOI shows much smaller relative deviations than the individual betas. The relative deviations are also typically smaller for the larger TOI (4.2in mortar).

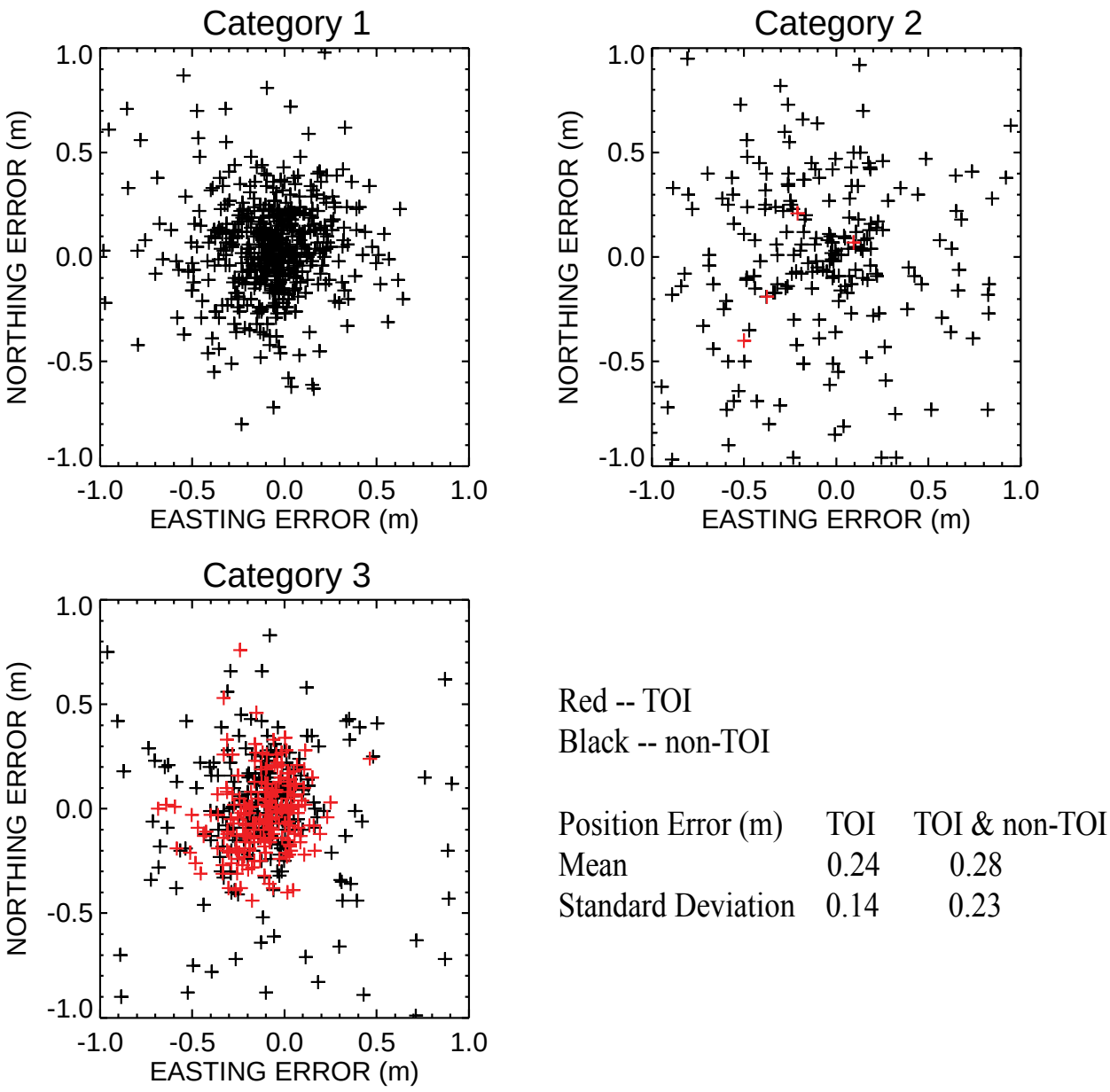


Figure 7-19. Differences between fitted and measured XY locations; EM61 Array.

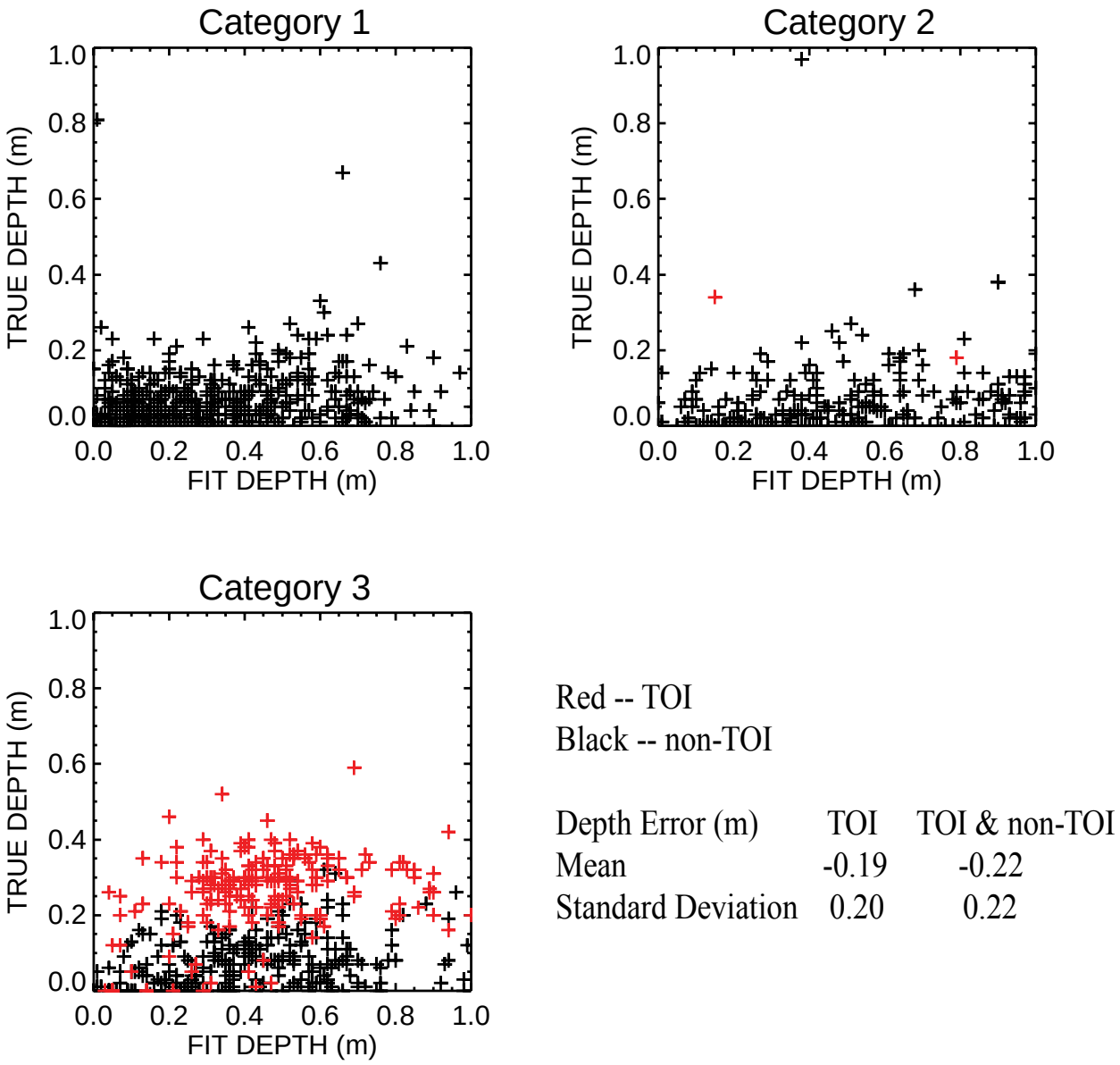


Figure 7-20. Fitted versus measured depth of burial; EM61 Array.

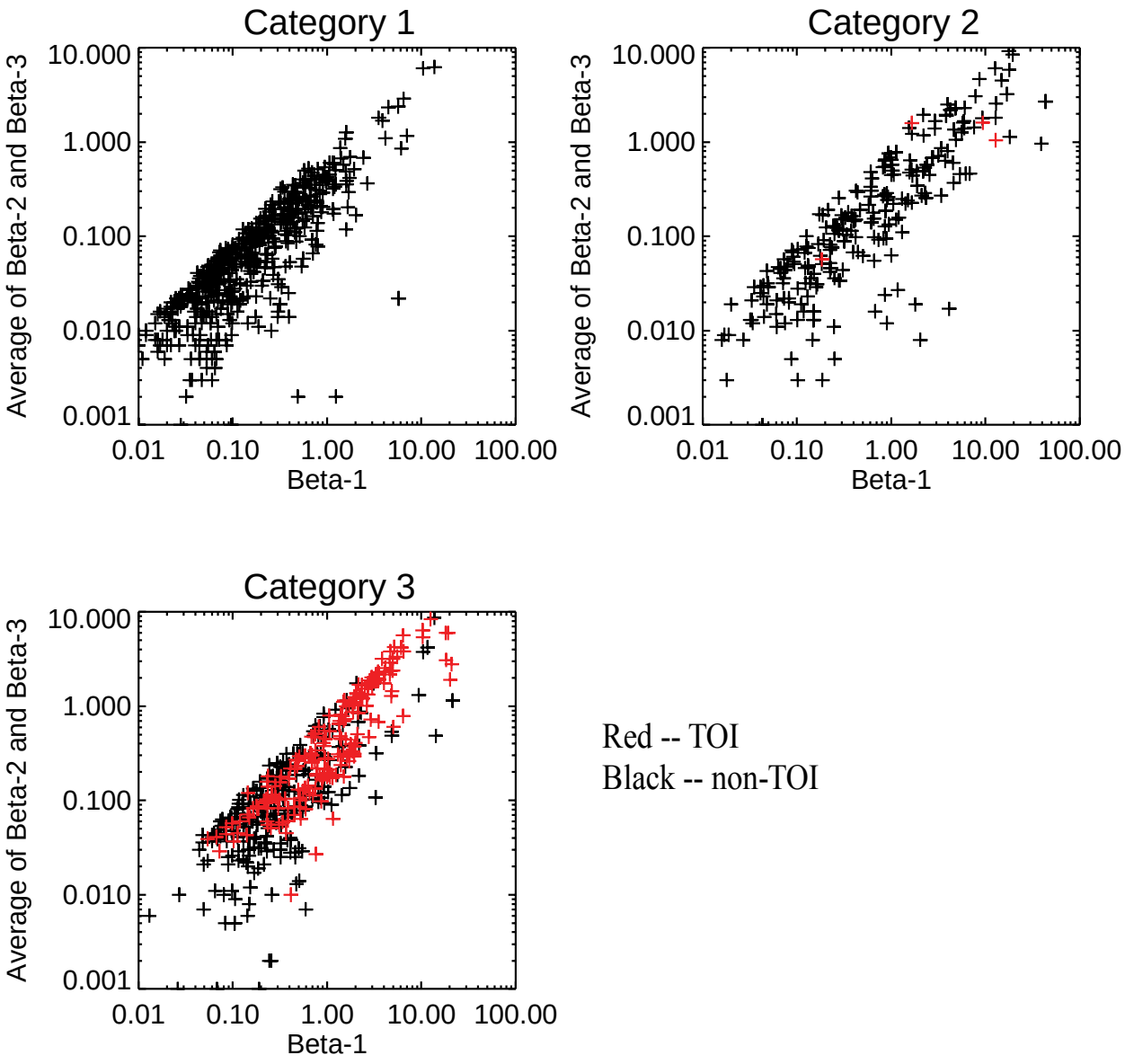


Figure 7-21. Beta 1 versus the average of Beta 2 and Beta 3; EM61 Array.

Table 7-6 Statistics of betas for the four main TOI, EM61 Array

Type	# of samples	Size		Beta 1		Beta 2		Beta 3	
		Mean	Std Dev	Mean	Std Dev	Mean	Std Dev	Mean	Std Dev
60mm	73	0.042	0.011	0.520	0.655	0.156	0.186	0.056	0.074
81mm	58	0.067	0.013	1.776	1.239	0.659	0.562	0.273	0.177
2.36in	19	0.052	0.007	0.893	0.392	0.195	0.076	0.073	0.048
4.2in	50	0.099	0.019	5.147	4.785	2.504	1.555	1.088	0.481

7.3.3 Failure Analyses

Using the thresholds adopted for this analysis, there were no false negatives for the EM61 array.

7.4 EM61 MSEMS

Data from the SCORR EM61 MSEMS is shown in Figure 7-22. The circles identify anomalies selected for the Test Set by the ESTCP Program Office while the squares represent the anomalies selected for training the classifiers.

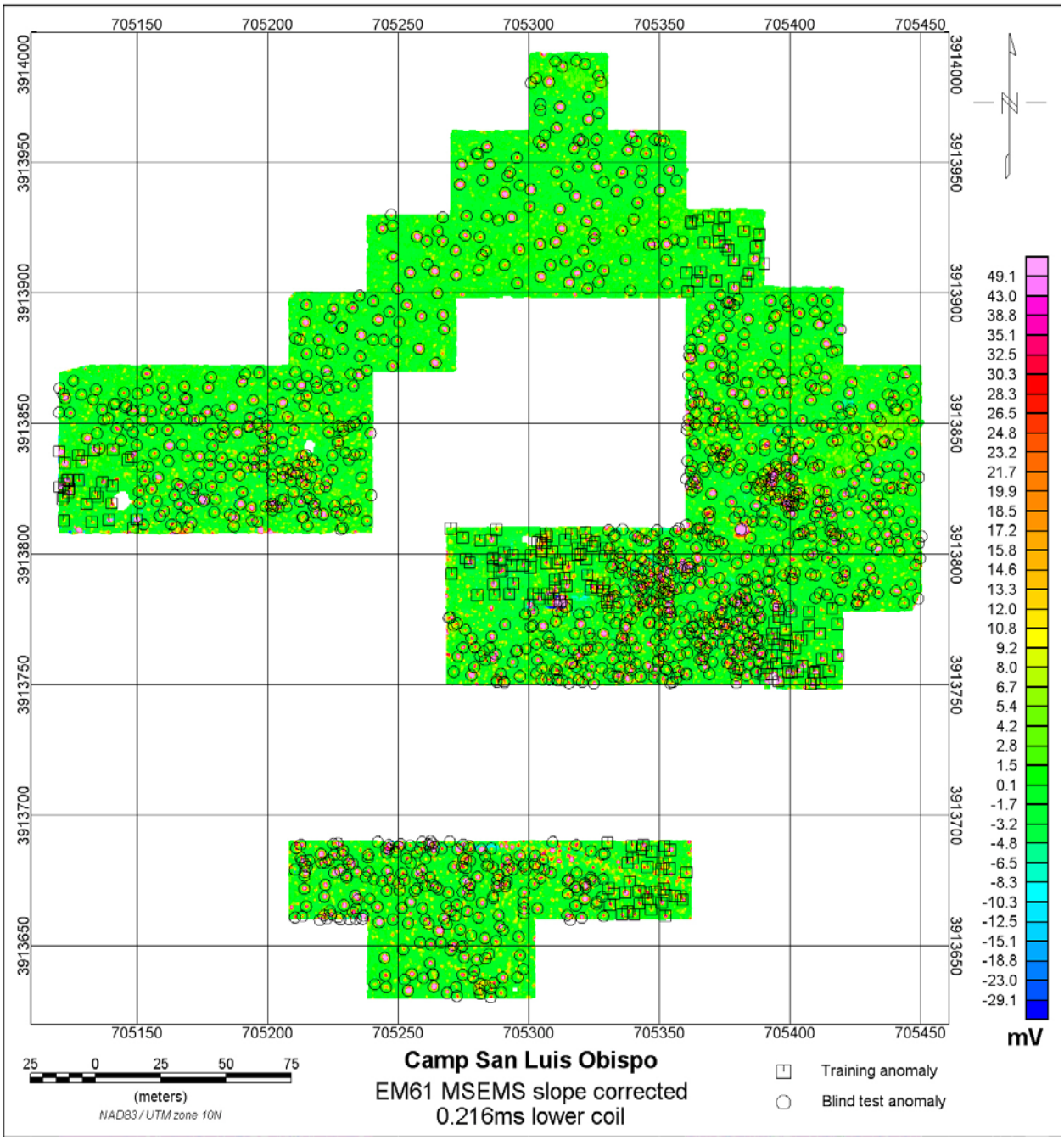


Figure 7-22. EM61 MSEMS data.

7.4.1 Performance Scores from IDA

Scoring performances for the EM61 MSEMS analysis are reported in Table 7-7 and shown graphically in Figure 7-23. A ROC chart is shown in Figure 7-24, where we plot the Percent of Munitions Dug versus the Number of Unnecessary Digs.

Using the thresholds adopted for this analysis, there were seven false negatives. They consisted of six 60mm mortars and one 2.36 inch rocket.

Table 7-7 Test Set Summary: EM61 MSEMS

Category	Cultural	Munition Debris	No Contact	Soil	UXO
1	20	709	9	0	7
2	3	156	11	0	2
3	9	179	3	0	196
4	0	0	0	0	0
TOTAL	**32**	**1044**	**23**	**0**	**205**

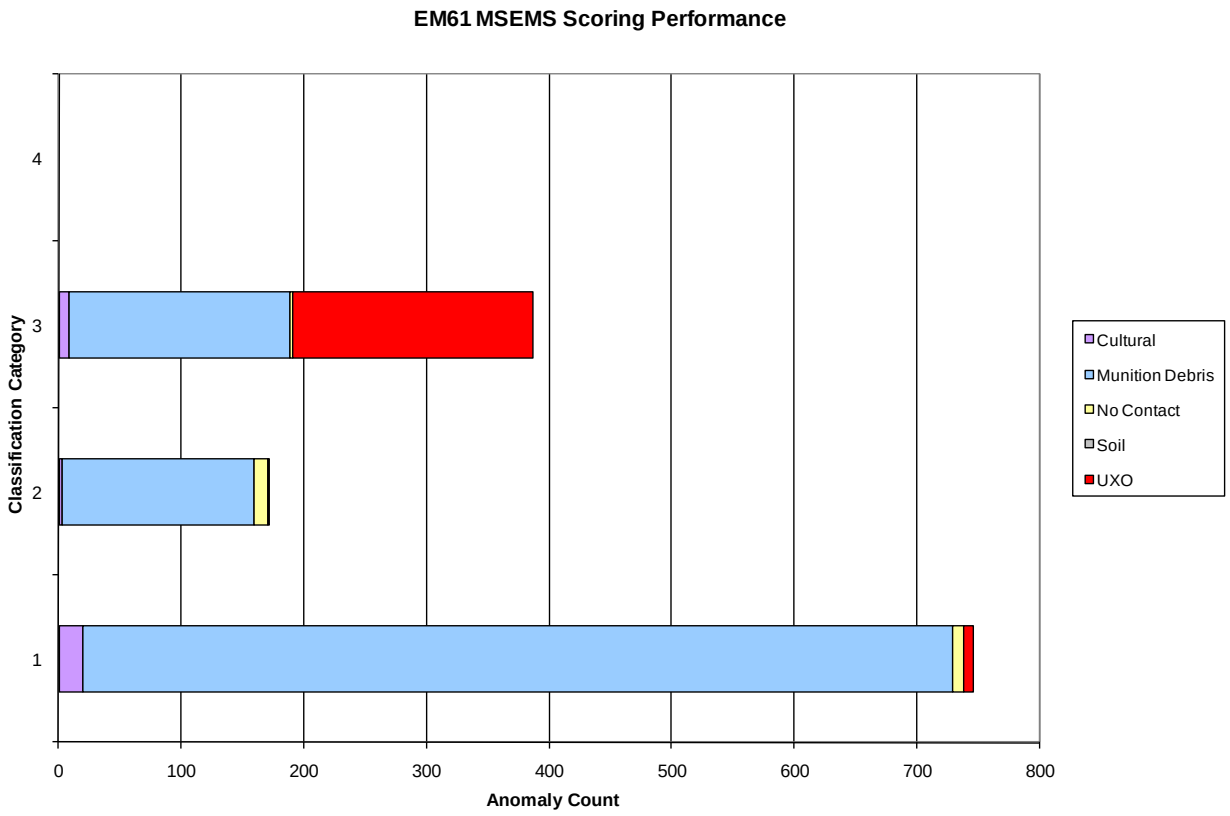


Figure 7-23. EM61 MSEMS performance as a function of classification category.

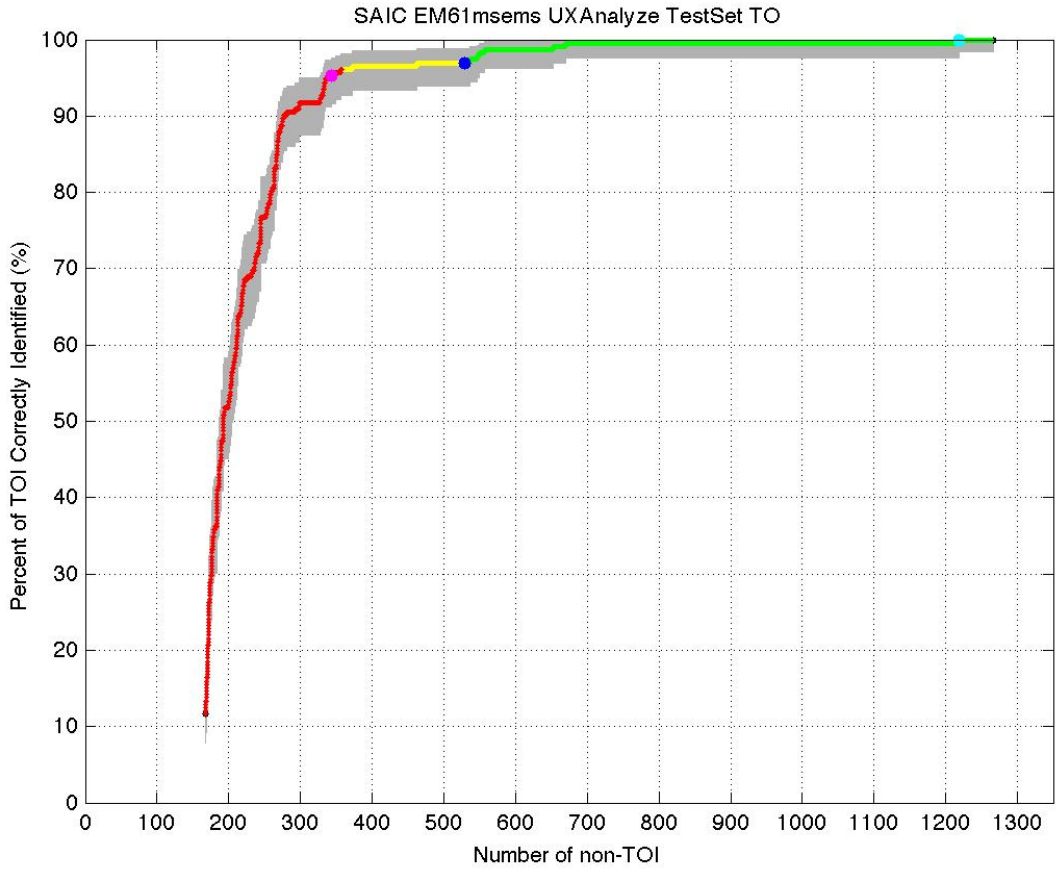


Figure 7-24. EM61 MSEMS ROC chart.

7.4.2 Characterization Plots

Figure 7-25 shows the difference between the fitted and measured XY locations for all category 1, 2 and 3 targets from the Test Set. The mean error for all TOI with an isolated or slightly overlapping signal was 0.22m with a standard deviation of .131m. If non-TOI are added to the population the mean error increases to 0.28m with a standard deviation of 0.21m.

Figure 7-26 shows the difference between the fitted and true depth for all category 1, 2 and 3 targets from the Test Set. The mean error for all TOI with an isolated or slightly overlapping signal was -0.13m with a standard deviation of .10m. If non-TOI are added to the population the mean error increases to -0.24m with a standard deviation of 0.16m.

Figure 7-27 and Table 7-8 show the inverted polarizabilities for all category 1, 2 and 3 targets from the Test Set. In general, the calculated size (sum of betas) of the main TOI shows much smaller relative deviations than the individual betas. The relative deviations are also typically smaller for the larger TOI (4.2in mortar).

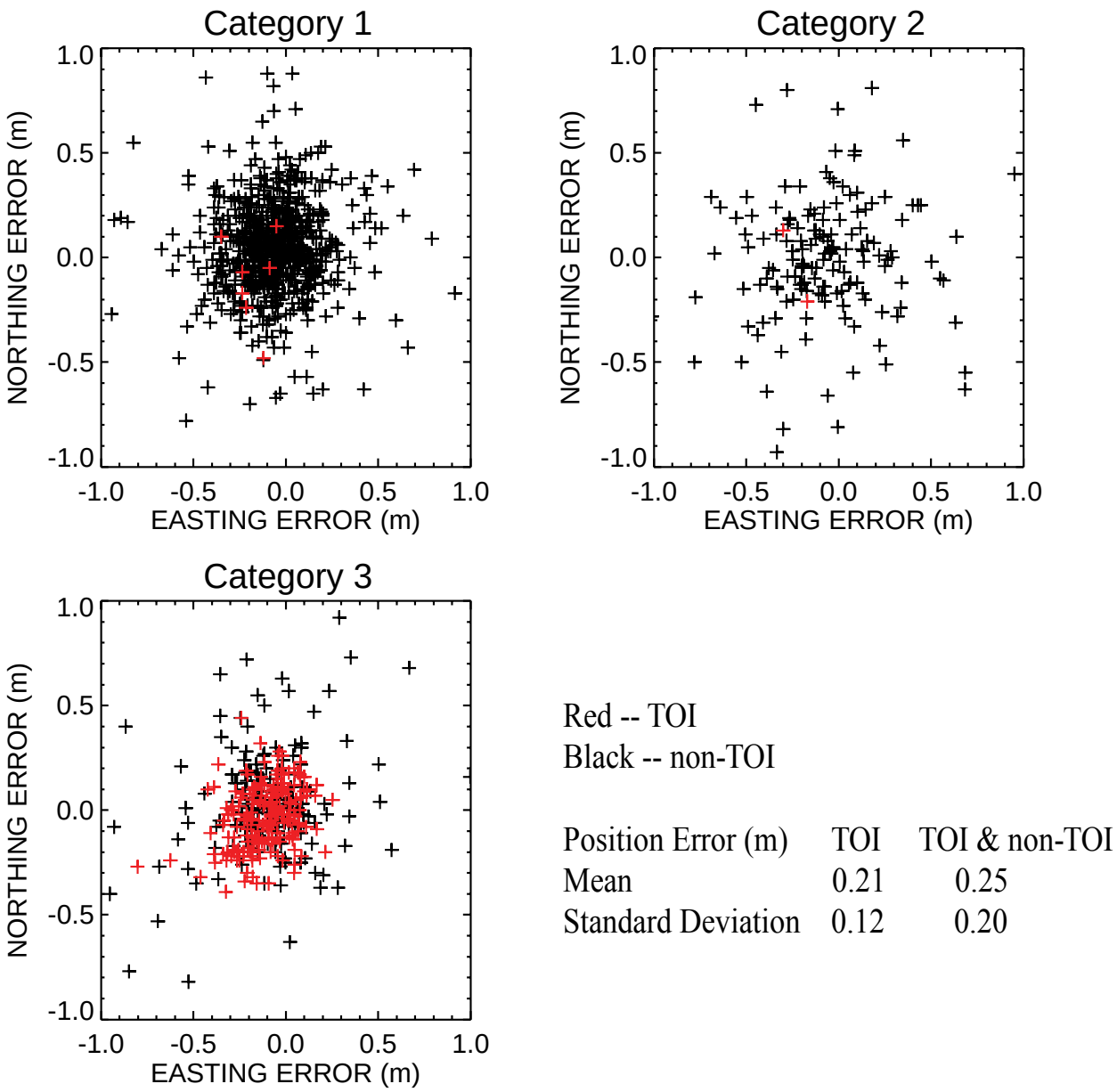


Figure 7-25. Differences between fitted and measured XY locations; EM61 MSEMS

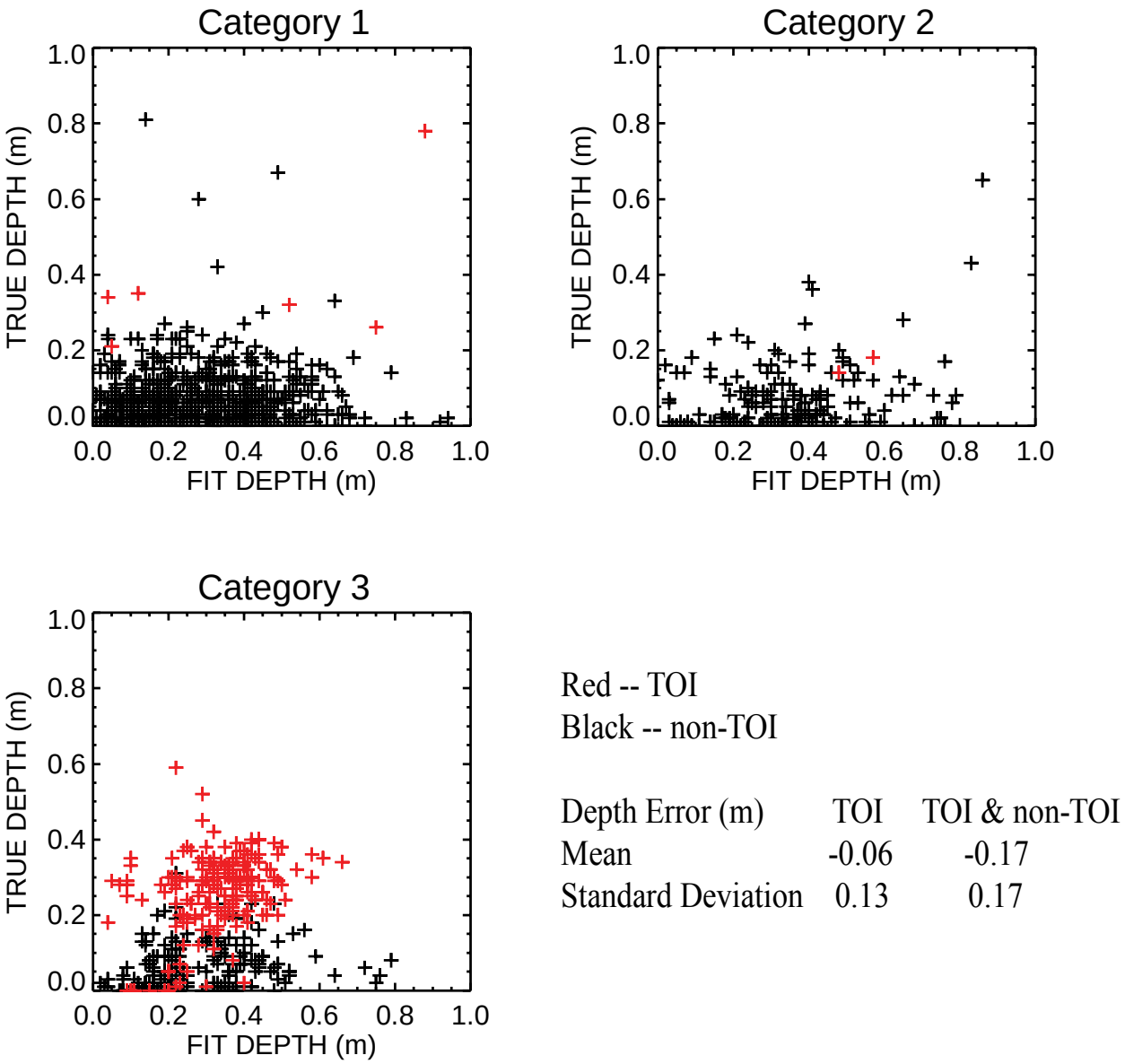


Figure 7-26. Fitted versus measured depth of burial; EM61 MSEMS

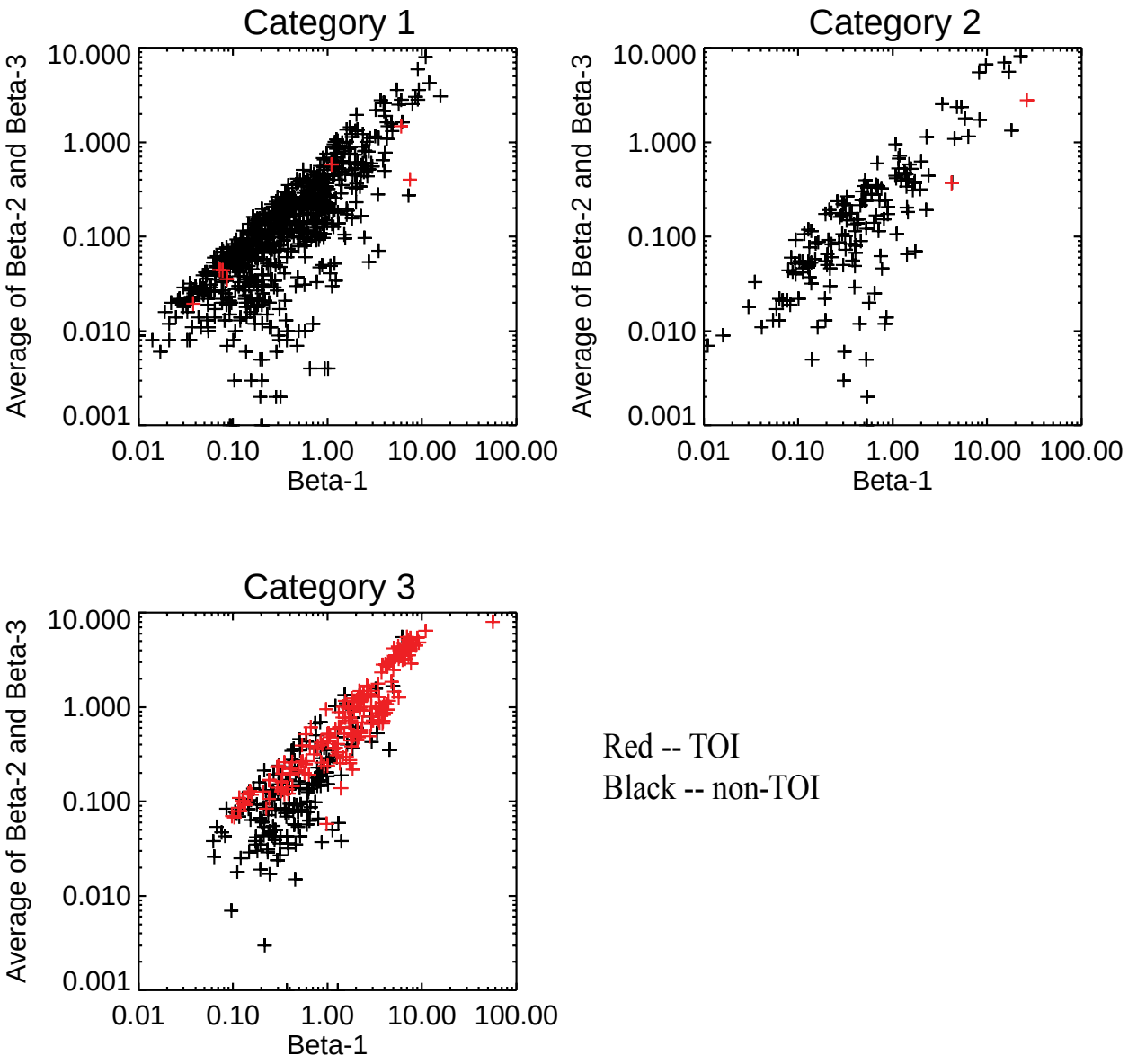


Figure 7-27. Beta 1 versus the average of Beta 2 and Beta 3; EM61 MSEMS

Table 7-8 Statistics of betas for the four main TOI, EM61 MSEMS

Type	# of samples	Size		Beta 1		Beta 2		Beta 3	
		Mean	Std Dev	Mean	Std Dev	Mean	Std Dev	Mean	Std Dev
60mm	73	0.048	0.013	0.673	0.558	0.266	0.210	0.106	0.107
81mm	56	0.080	0.011	2.541	1.183	1.024	0.342	0.655	0.309
2.36in	18	0.072	0.013	2.409	1.272	0.574	0.365	0.325	0.270
4.2in	51	0.115	0.011	6.092	1.610	3.785	0.965	2.659	0.768

7.4.3 Failure Analyses

Four of the seven failures were due to an aggressive choice of the "dig/no dig" threshold which placed anomalies 675, 775, 1285 and 1441 into the no dig category. All these anomalies were 60mm mortars and had a decision metric of greater than 0.3. The Category 1-2 threshold used for the EM61 MSEMS analysis was 0.38. In comparison, the EM61 MK2 cart analysis used a threshold of 0.1. If the 0.1 threshold was used, all the above mentioned anomalies would have been placed in Category 3. In addition to the aforementioned anomalies, anomaly 314 (60mm mortar) would also have been placed into Category 3. Figure 7-28 shows the EM61 MSEMS ROC curve using the 0.1 threshold. The remaining failures are anomaly 1444 and 444.

Anomaly #1444 was a seeded 60mm body with a decision metric of 0.099 which was slightly below our Category 1-2 threshold of 0.1. This anomaly was also misclassified by the EM61 slope and non SCORR data. Figure 7-29 shows the EM61 MSEMS measured and modeled data for anomaly 1444. As with the EM61 MK2 cart data, close examination of the EM61 MSEMS data potentially shows some small sources near the anomaly (indicated by the circles) that may have affected the data sufficiently enough to cause the failure.

The final problematic target was Master ID 444, a 2.36 inch rocket at a depth of 78cm. The burial depth was much deeper than the maximum depth of 45cm that was used at this site. There were multiple ground truth objects (Figure 7-30) and several other anomalies that overlapped with this anomaly which caused problems with the analysis. Figure 7-31 shows the measured data (top left) and modeled data for this anomaly as well as the locations of other target declarations. There were multiple targets picked near this anomaly and the data clearly shows several overlapping signatures. Master ID 435 shows a distinct anomaly even though the ground truth indicated a "no contact". Anomaly 444 was classified a Category 1 target based on a low decay ratio. The profile at the bottom of Figure 7-31 shows 2 peaks over the picked location. The 2 peaks have distinct decay ratios with the peak on the left exhibiting a decay similar to a TOI while the peak on the right shows a much smaller ratio indicative of thin walled object. We presume the low ratio was caused by multiple source objects that when combined resulted in a decay ratio less than a TOI. Retrospectively, this target should have been classified as a Category 2 target due to the overlapping signals.

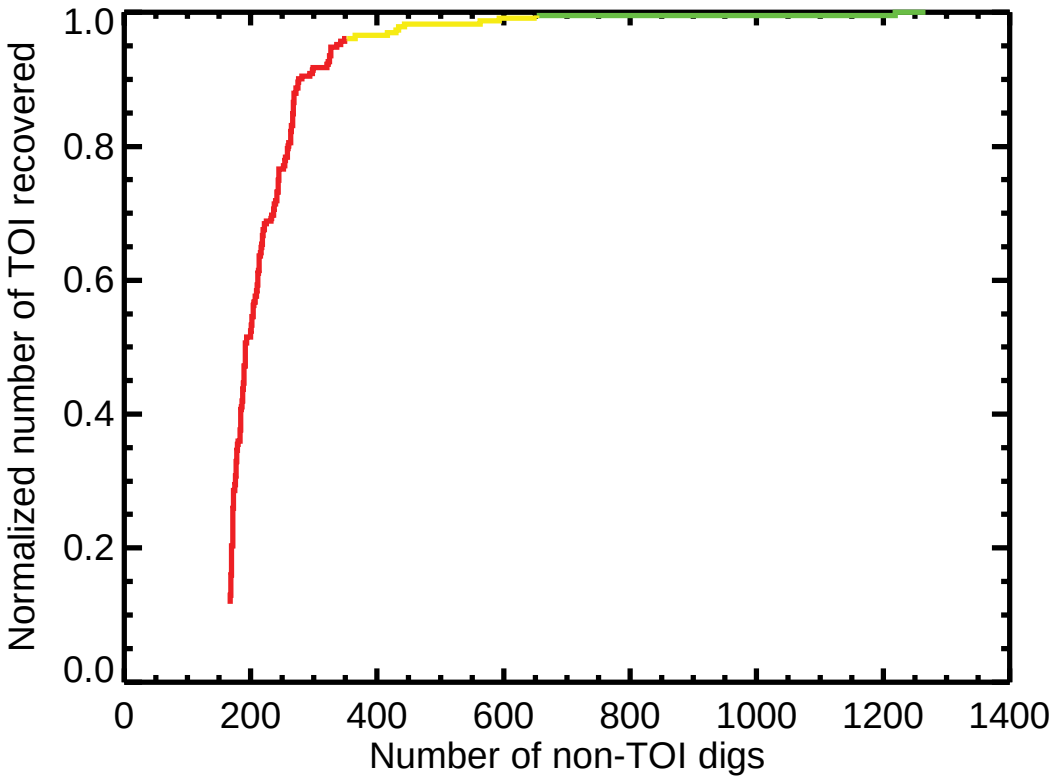
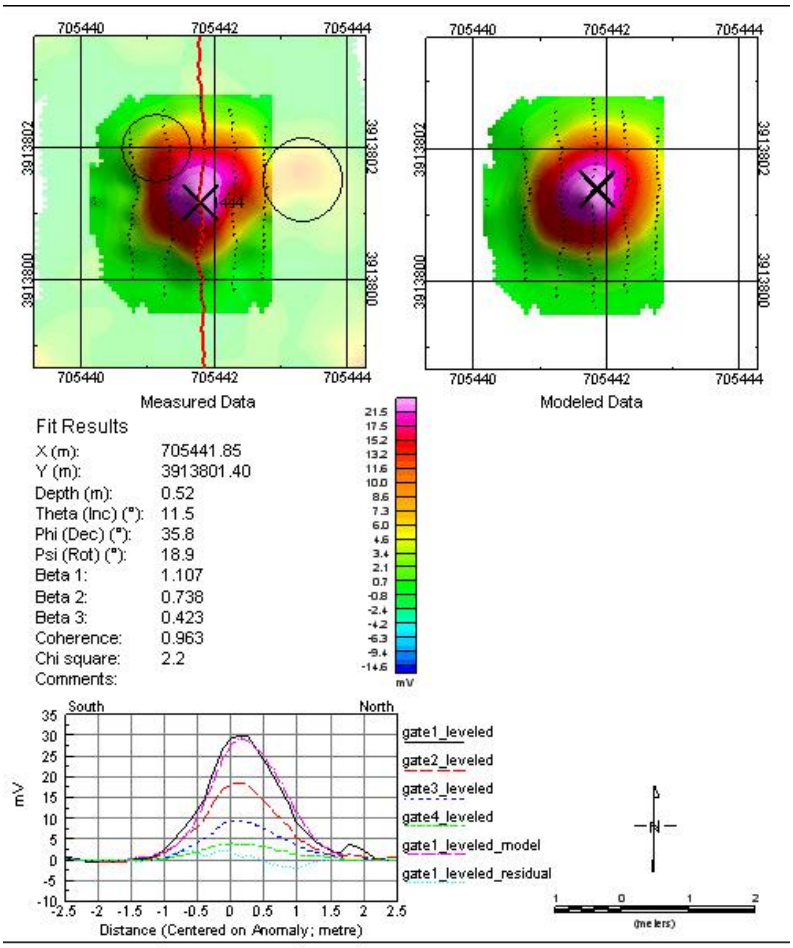


Figure 7-28. EM61 MSEMS ROC curve using a more conservative Category 1-2 threshold of 0.1.



SLO - EM61 MSEMS - Anomaly 1444

Figure 7-29. Anomaly plot showing measured data, inverted features and forward model for anomaly 1444 which was ranked just below the Category 1-2 boundary.

Figure 7-30. Photographs of objects excavated at Target 444. The whiteboard on the left inadvertently shows target id 435.

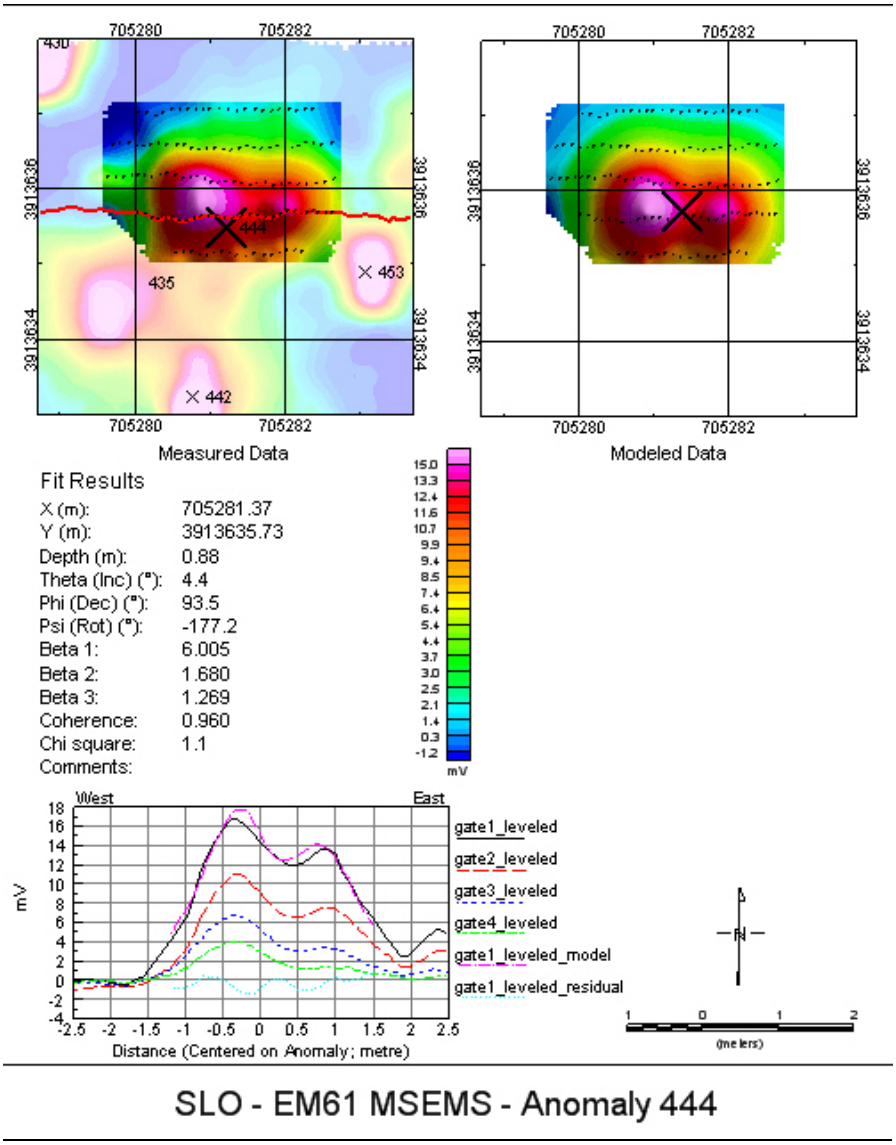


Figure 7-31. Anomaly plot showing measured data, inverted features and forward model for anomaly 444.

7.5 TEMTADS

Data from the TEMTADS is shown in Figure 7-32. The circles identify anomalies selected for the Test Set by the ESTCP Program Office while the squares represent the anomalies selected for training the classifiers.

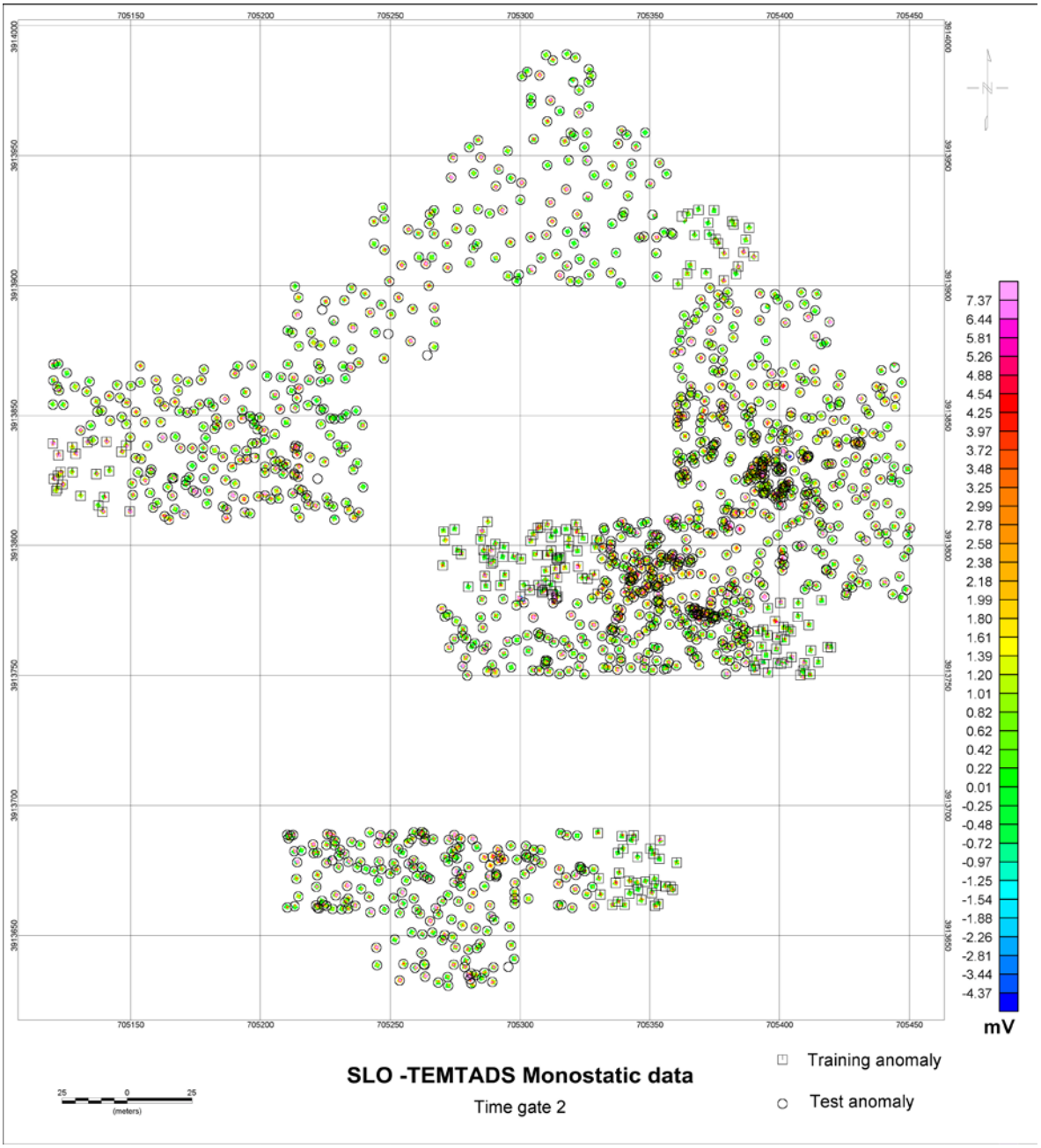


Figure 7-32. TEMTADS data.

7.5.1 Performance Scores from IDA

Scoring performances for the 2 and 3 criteria TEMTADS analysis are reported in Table 7-9 and Table 7-10 and shown graphically in Figure 7-33 and Figure 7-35. Their respective ROC chart are shown in Figure 7-34 and Figure 7-36, where we plot the Percent of Munitions Dug versus the Number of Unnecessary Digs.

Using the thresholds adopted for this analysis, there were 3 false negatives for the 2 criteria analysis while the 3 criteria analysis had 7 false negatives.

Table 7-9 Test Set Summary: TEMTADS – 2 criteria

Category	Cultural	Munition Debris	No Contact	Soil	UXO
1	29	706	15	0	3
2	5	252	23	1	14
3	2	35	1	0	187
4	1	5	1	0	2
TOTAL	**37**	**998**	**40**	**1**	**206**

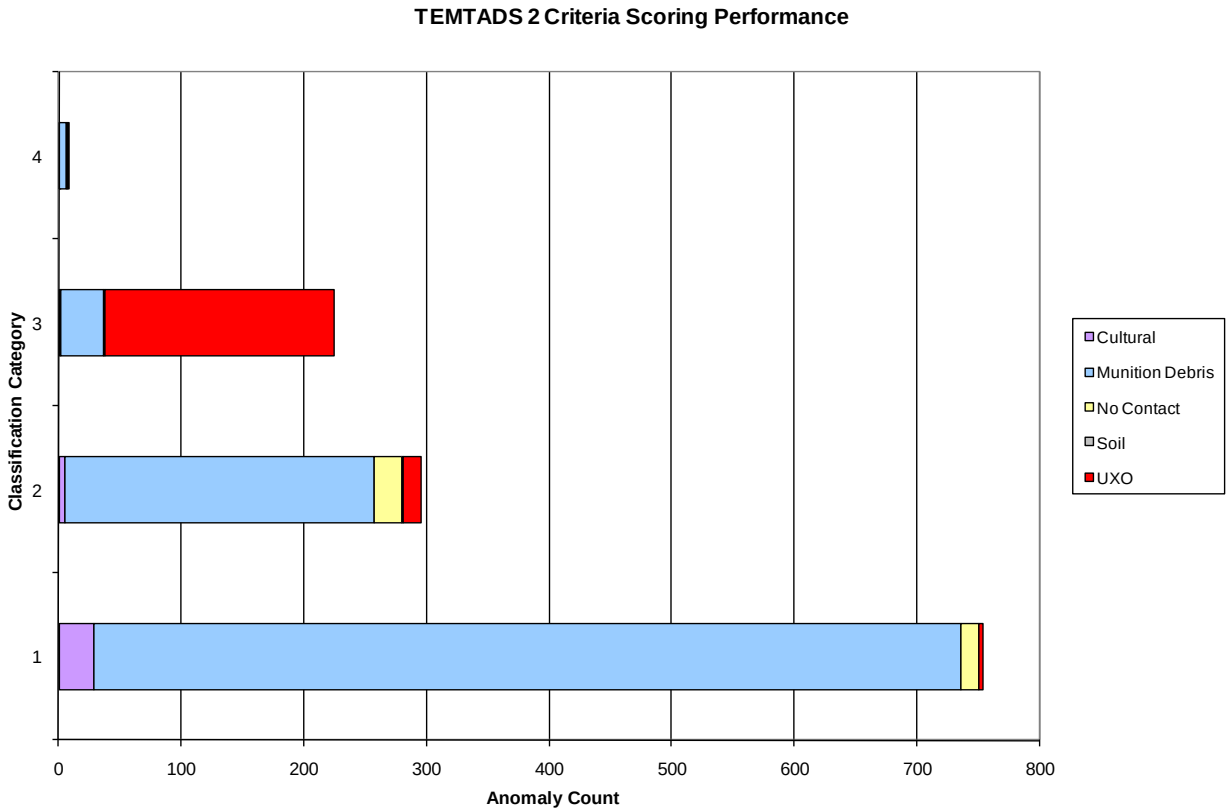


Figure 7-33. TEMTADS 2 criteria performance as a function of classification category.

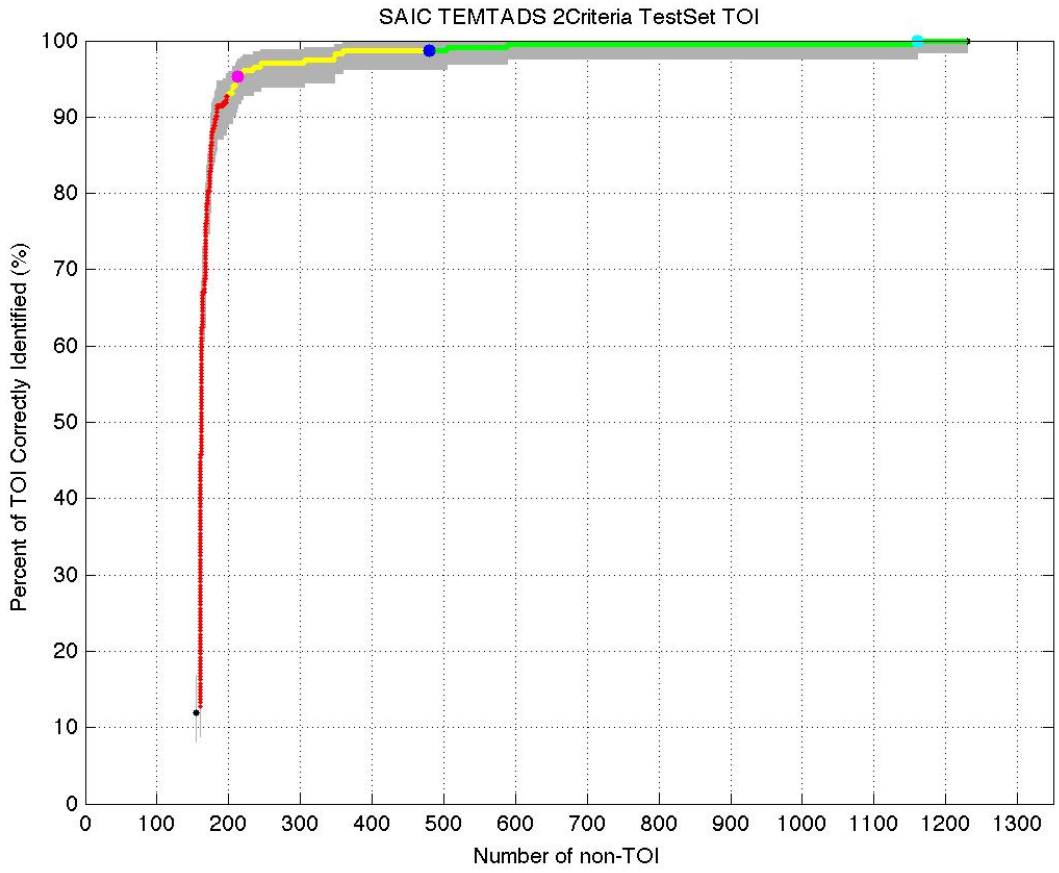


Figure 7-34. TEMTADS 2 criteria ROC chart.

Table 7-10 Test Set Summary: TEMTADS – 3 criteria

Category	Cultural	Munition Debris	No Contact	Soil	UXO
1	31	698	15	0	7
2	5	271	24	1	13
3	0	23	1	0	184
4	1	6	0	0	2
TOTAL	**37**	**998**	**40**	**1**	**206**

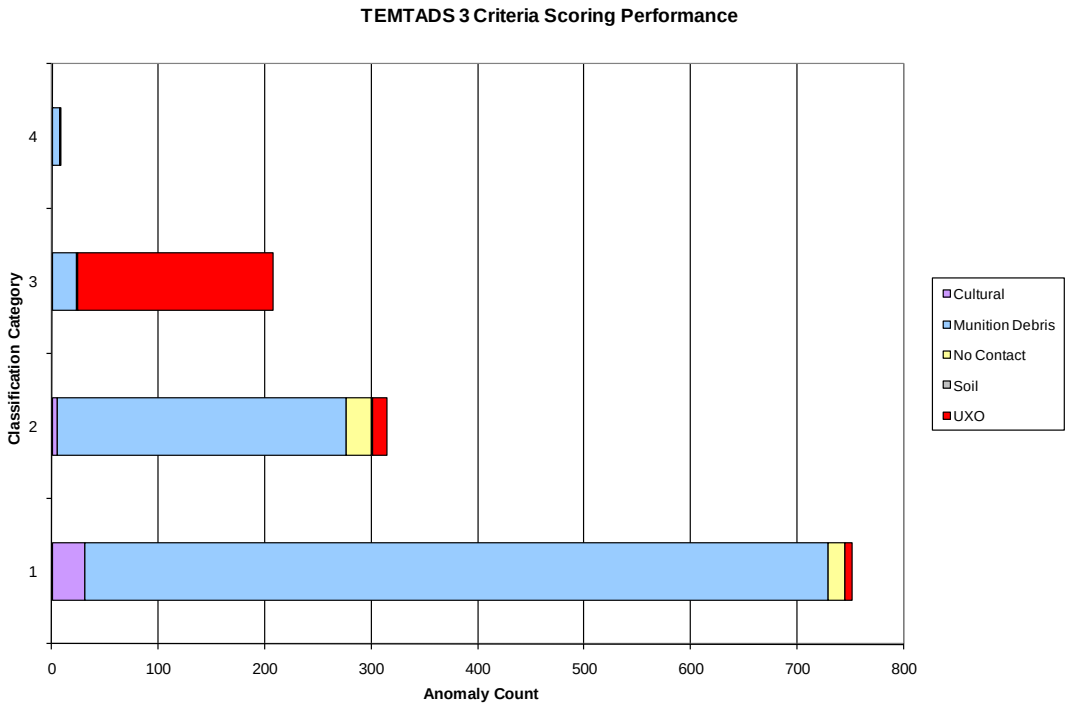


Figure 7-35. TEMTADS 3 criteria performance as a function of classification category.

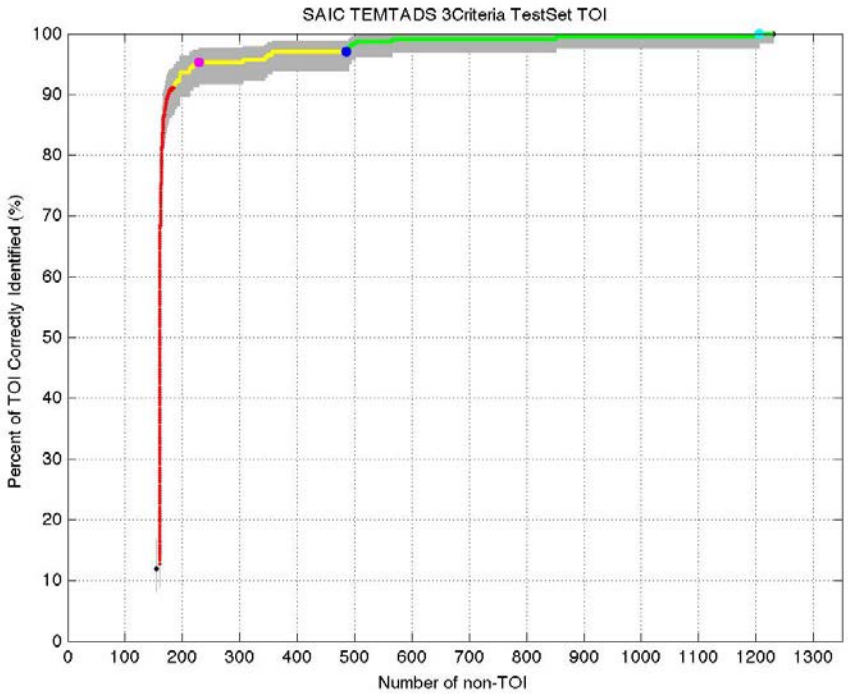


Figure 7-36. TEMTADS 3 criteria ROC chart.

7.5.2 Characterization Plots

Figure 7-37 shows the difference between the fitted and measured XY locations for all category 1, 2 and 3 targets from the Test Set. The mean error for all TOI with an isolated or slightly overlapping signal was 0.16m with a standard deviation of .14m. If non-TOI are added to the population the mean error increases to 0.22m with a standard deviation of 0.26m.

Figure 7-38 shows the difference between the fitted and true depth for all category 1, 2 and 3 targets from the Test Set. The mean error for all TOI with an isolated or slightly overlapping signal was -0.01m with a standard deviation of .09m. If non-TOI are added to the population the mean error is also -0.01m with a standard deviation of 0.12m.

In Figure 7-39, we plot the median and 1 standard deviation error bars for our derived β's for all UXO/TOI in the Test Set, divided into the four main types. The remarkable consistency shown in these plots validates our direct library matching technique. The 60mm mortars show the most variation overall, due to the presence of a large number of objects without the tail assembly. The high fidelity data obtained by the TEMTADS allows the use of shape to characterize the targets instead of only relying on size which was the case for the EM61 sensors.

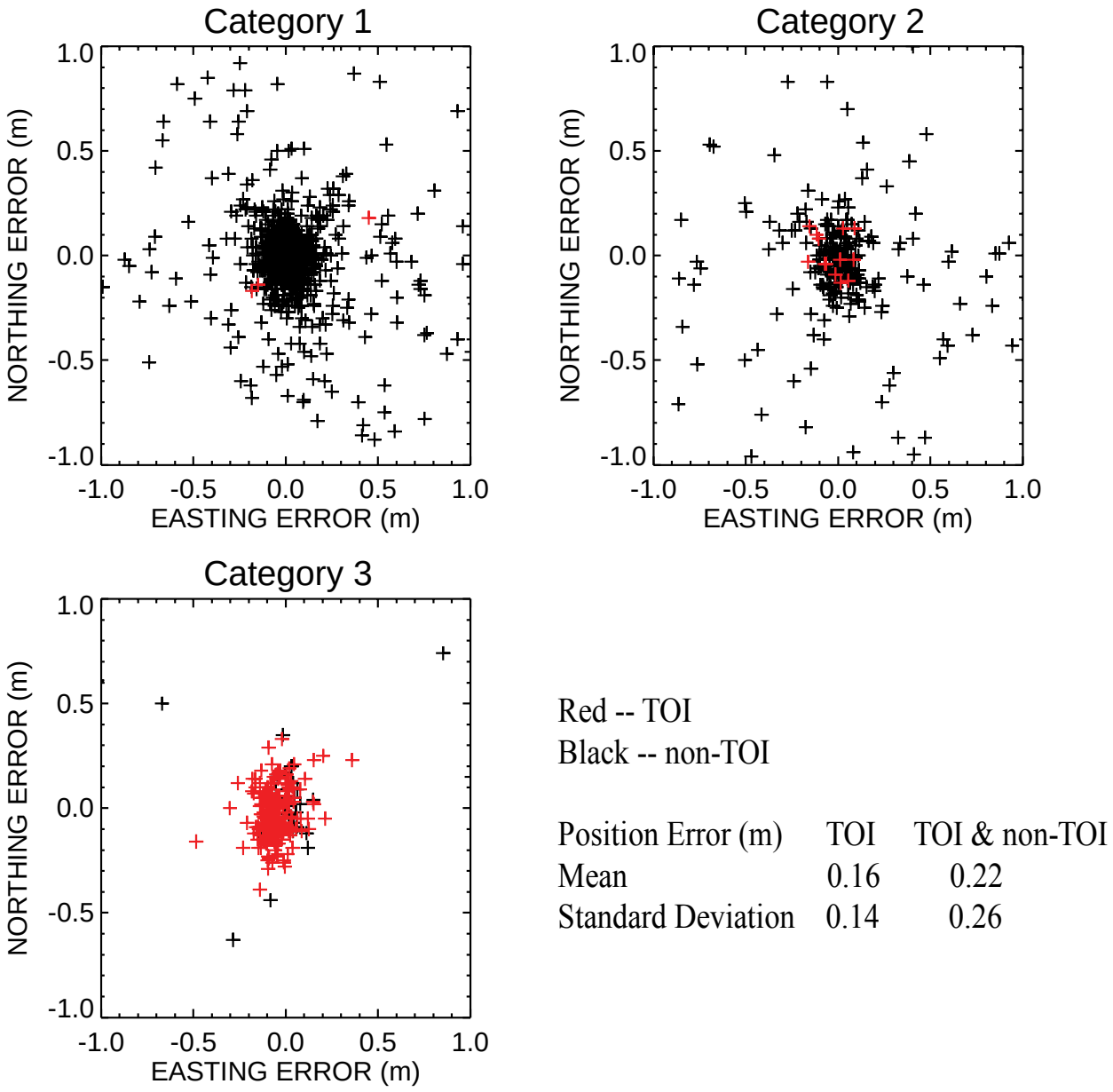


Figure 7-37. Differences between fitted and measured XY locations; TEMTADS 2 criteria analysis

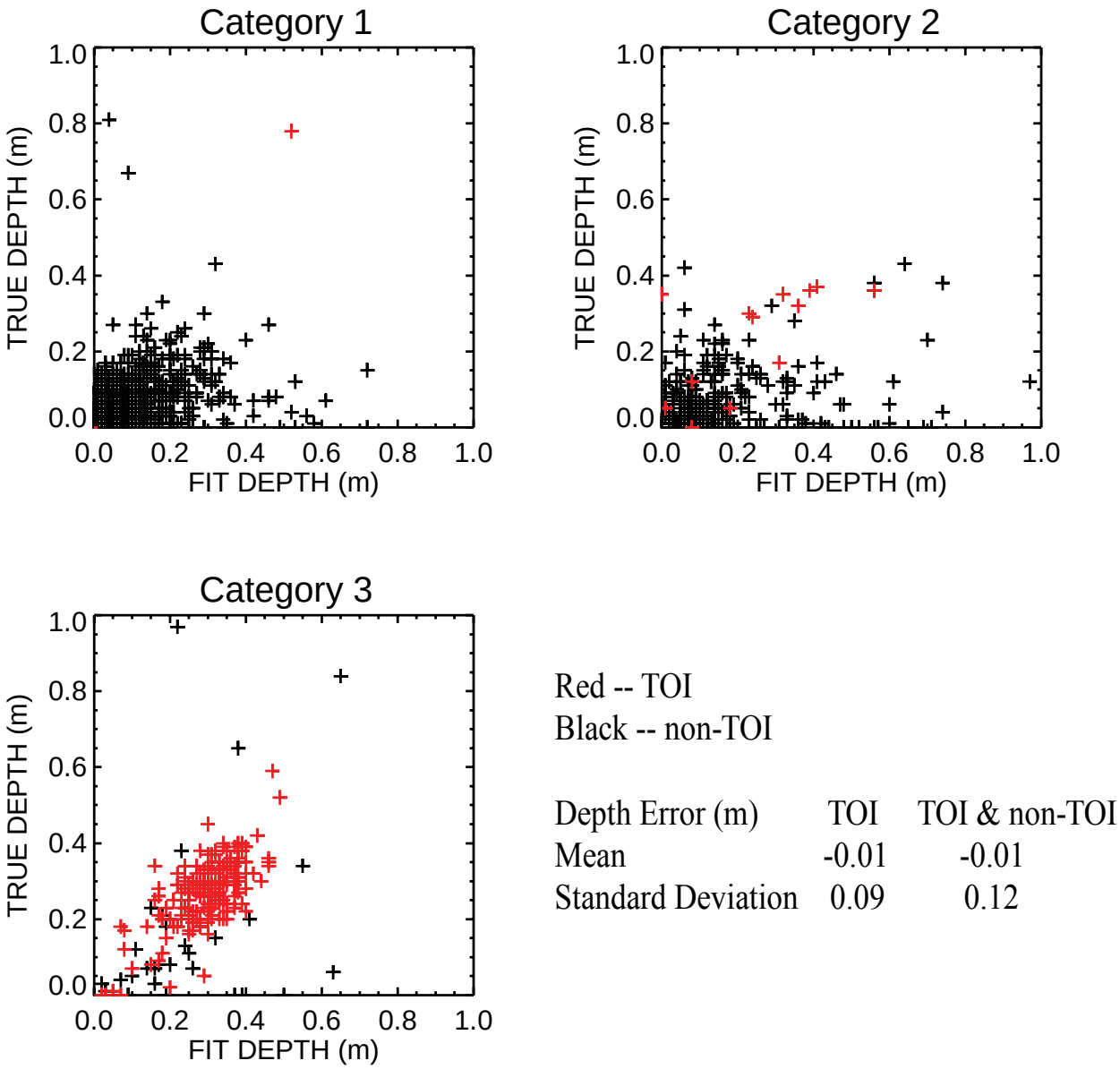


Figure 7-38. Fitted versus measured depth of burial; TEMTADS 2 criteria analysis

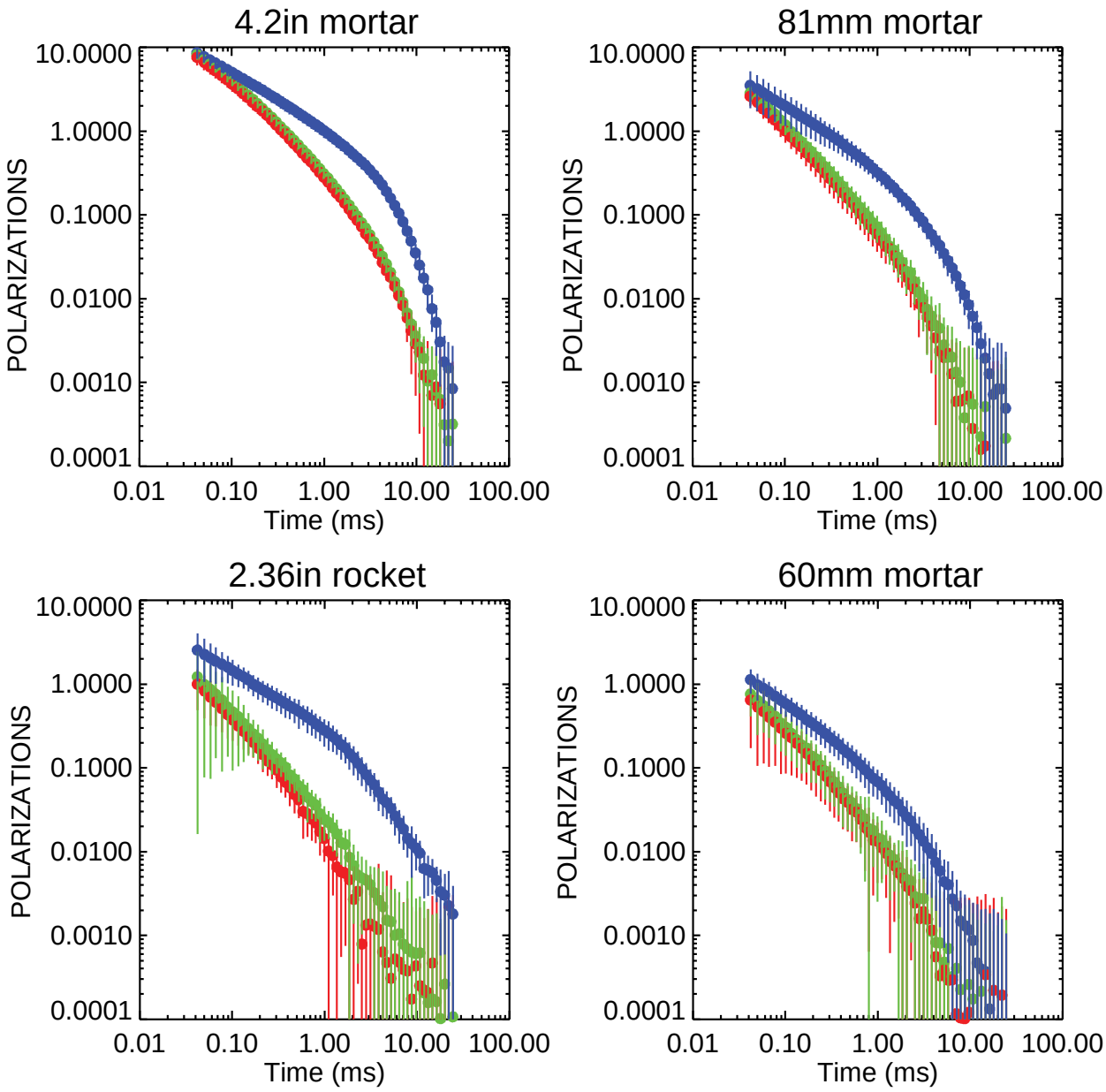


Figure 7-39. Median and ±1σ errors of the four main UXO for the TEMTADS test set anomalies.

7.5.3 Failure Analyses

In Table 7-11, we give the Master IDs of our false negatives for both our 3 and 2 criteria methods. The 3 criteria method contained all 3 false negatives that the 2 criteria method has, along with 4 others.

Table 7-11 TEMTADS' False Negatives

Master ID	3 Criteria/2 Criteria Method
16	3
65	3
241	Both
249	3
372	3
444	Both
711	Both

All four of the UXO/TOI that were missed only by the 3 criteria method are 60mm's with low SNR. Two of them had peak signals close to the 2mV/Amp cutoff that was used to classify targets as Category 2. For these cases, the two larger betas were matched reasonably well, but the smallest beta was not. This is the reason the 2 criteria analysis correctly classified the targets as TOI while the 3 criteria did not. This is illustrated for anomaly 16 in Figure 7-40 where the red lines represent the library data and the blue lines plot the inverted polarizations for the 3 betas.

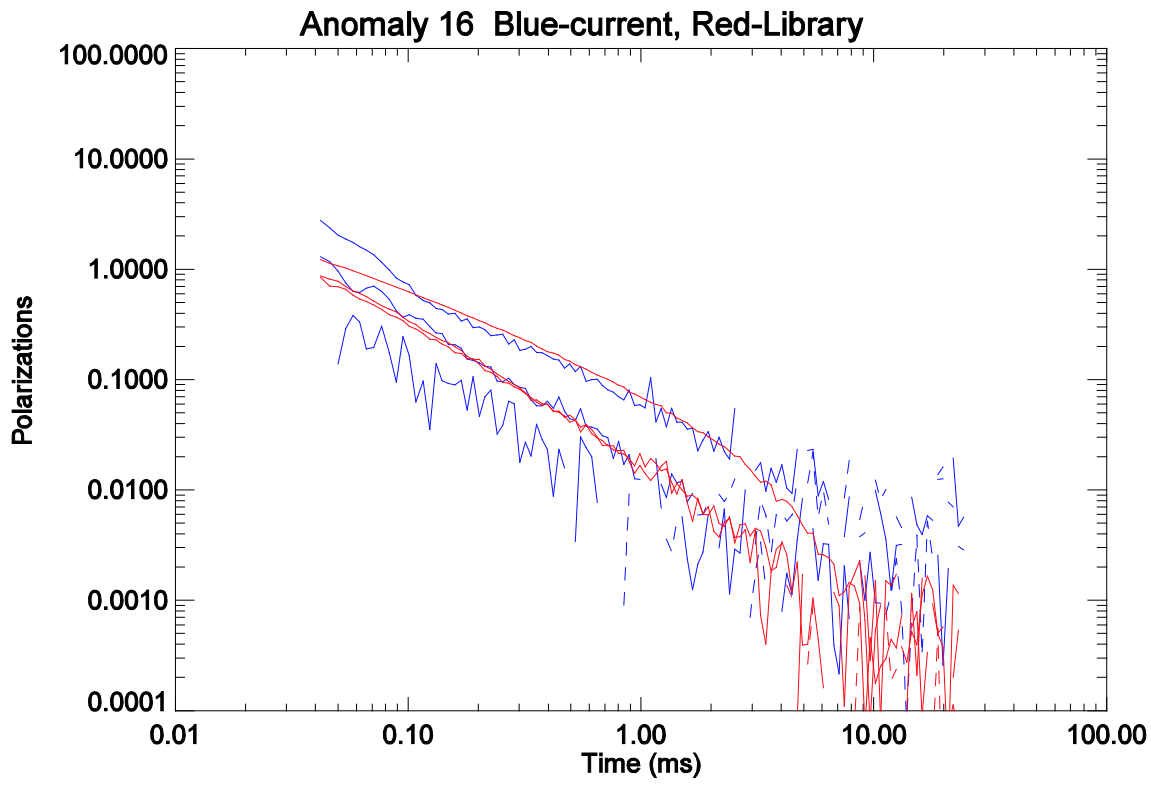


Figure 7-40. TEMTADS library match to anomaly 16

Of the 3 false negatives in common to both methods, anomalies 241 and 444 both contain 2.36 inch rockets along with a half-dozen pieces of frag, as shown previously in Figure 7-30 and Figure 7-41 below. The presence of multiple sources was the cause of these two failures.

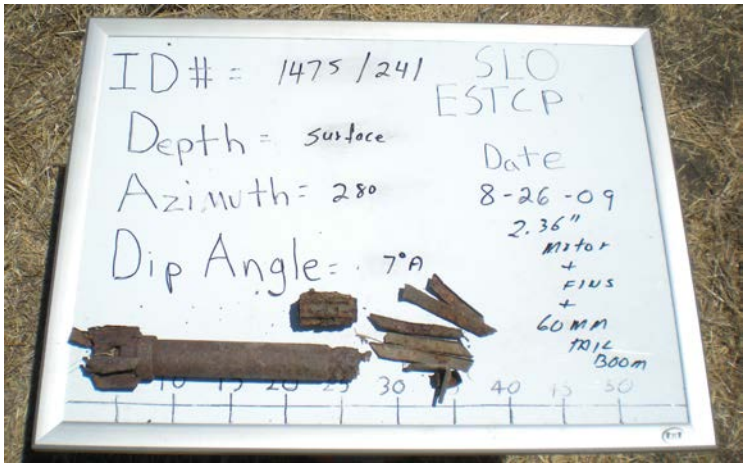


Figure 7-41. Photograph of objects excavated at Target 241. In addition to a 2.36" mortar, a number of fragments were found.

A contour plot of the remaining anomaly, 711, is shown in Figure 7-42. The dig location is marked by a circle at the upper edge of the map, some 0.65m away from the predicted location, as noted by the "X". As a result, the target lies on the edge of our array, and the measured signal suggests we do not have the peak. This is therefore a sampling error. Steps were taken on site to examine the contour plots to see if any anomalies should be remeasured but in this case, the combination of confusion due to the multiple peaks, and the relatively low signal, led to our decision not to obtain a second measurement.

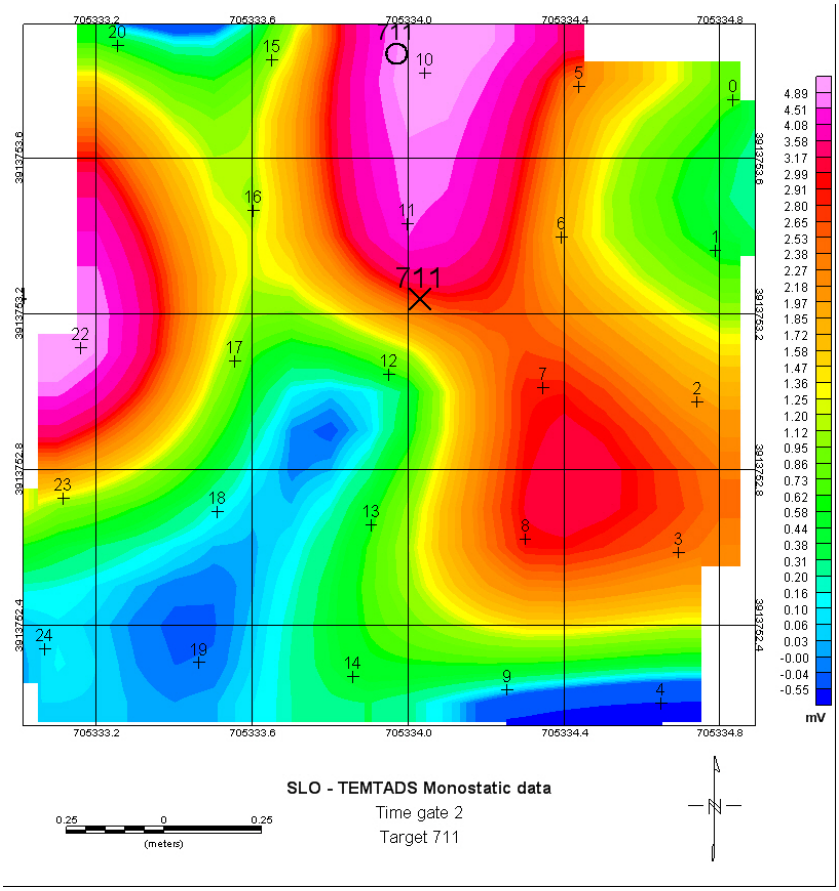


Figure 7-42. Color contour plot of the monostatic TEMTADS data for Target 711.

7.6 METAL MAPPER

Anomaly locations for the Metal Mapper are shown in Figure 7-43. As mentioned earlier the anomaly locations were selected from the dynamic Metal Mapper survey and not by the ESCTP Program Office. The black circles identify anomalies that were excavated and scored by the ESTCP Program Office while the squares represent the anomalies selected for training the

classifiers. The red circles show anomalies that were interrogated with the Metal Mapper but were not scored because they were not excavated.

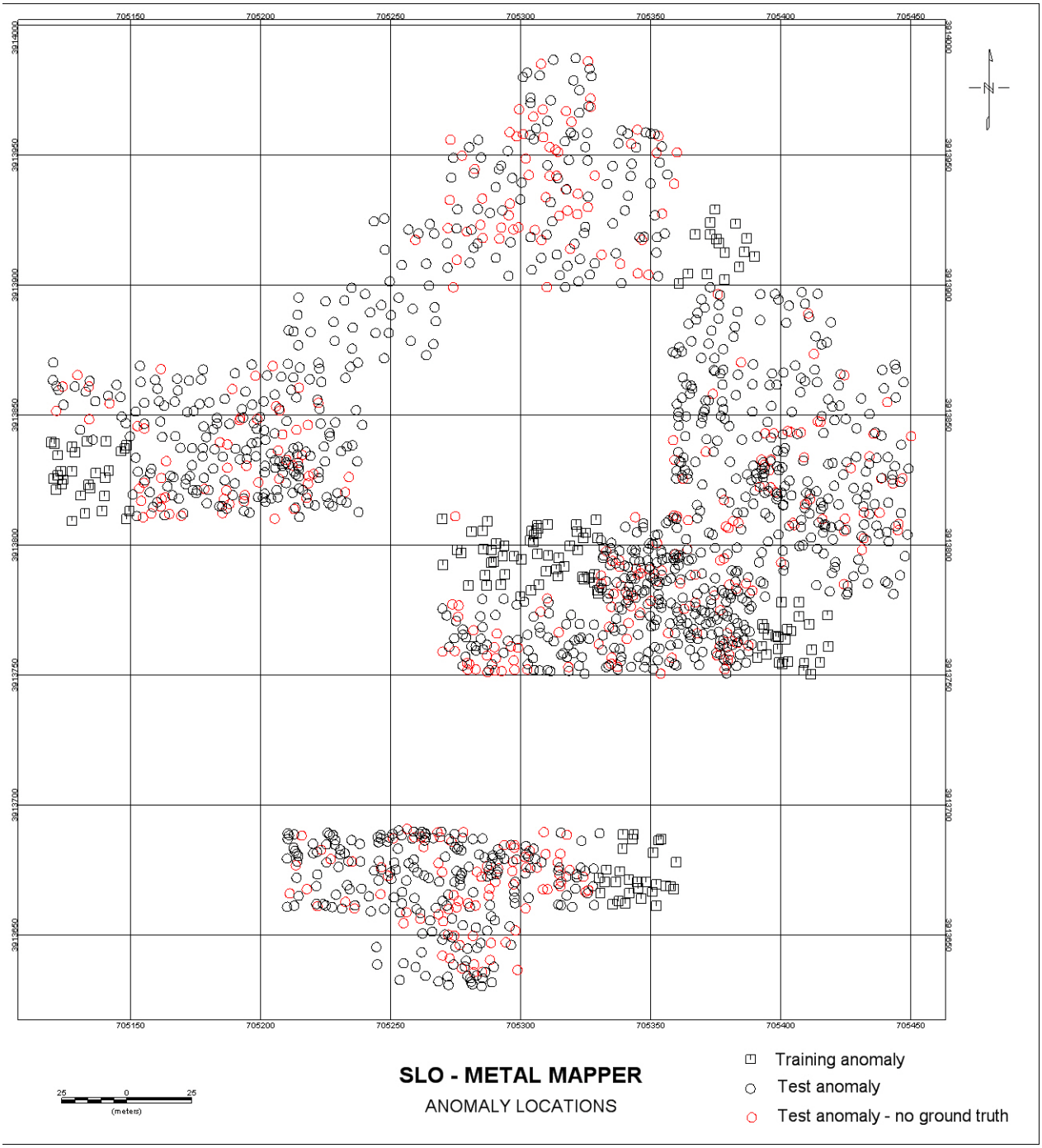


Figure 7-43. Metal Mapper training and testing anomaly locations.

7.6.1 Performance Scores from IDA

Scoring performances for the 2 and 3 criteria Metal Mapper analysis are reported in Table 7-12 and Table 7-13 and shown graphically in Figure 7-44 and Figure 7-46. Their respective ROC chart are shown in Figure 7-45 and Figure 7-47, where we plot the Percent of Munitions Dug versus the Number of Unnecessary Digs.

Using the thresholds adopted for this analysis, there were 5 false negatives for the 2 criteria analysis while the 3 criteria analysis had 4 false negatives.

Table 7-12 Test Set Summary: Metal Mapper – 2 criteria

Category	Cultural	Munition Debris	No Contact	Soil	Misc Clutter	UXO
1	19	594	10	0	23	5
2	2	162	4	0	2	4
3	2	37	2	0	1	195
4	0	1	0	0	0	0
TOTAL	**23**	**794**	**16**	**0**	**26**	**204**

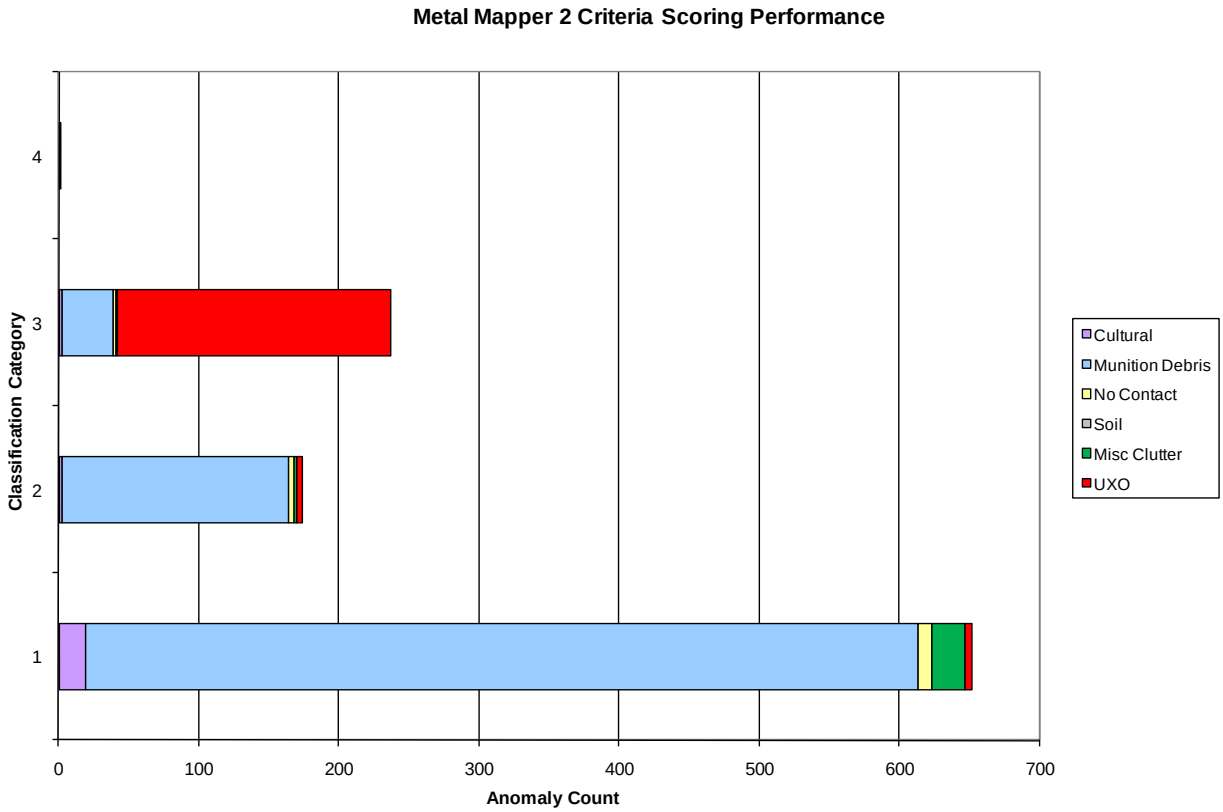


Figure 7-44. Metal Mapper 2 criteria performance as a function of classification category.

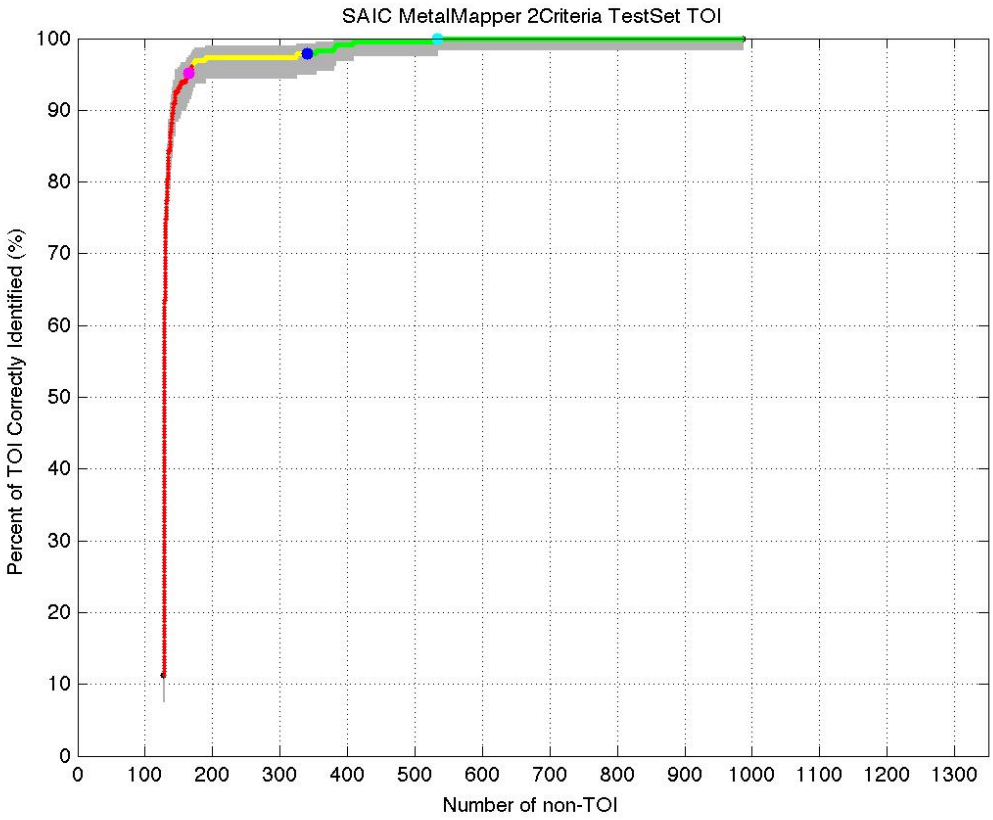


Figure 7-45. Metal Mapper 2 criteria ROC chart.

Table 7-13 Test Set Summary: Metal Mapper – 3 criteria

Category	Cultural	Munition Debris	No Contact	Soil	Misc Clutter	UXO
1	20	582	11	0	23	4
2	2	180	4	0	3	3
3	1	31	1	0	0	197
4	0	1	0	0	0	0
TOTAL	**23**	**794**	**16**	**0**	**26**	**204**

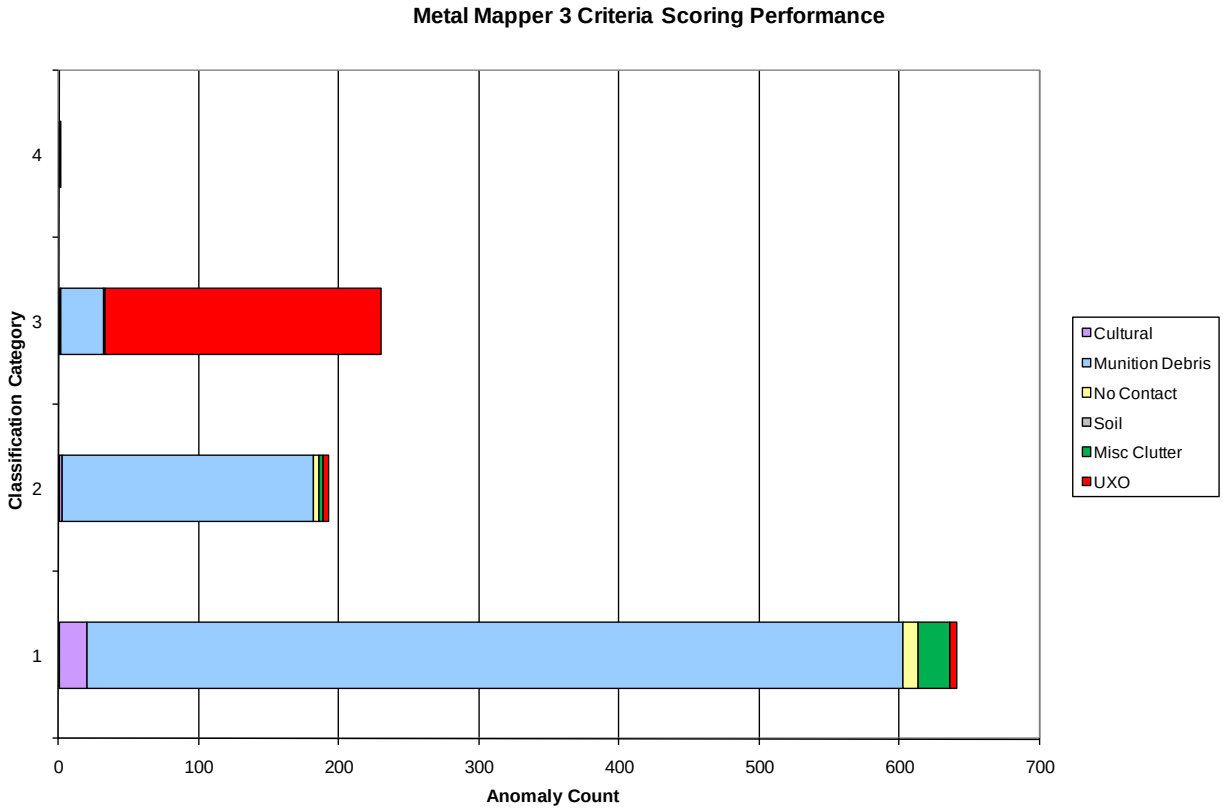


Figure 7-46. Metal Mapper 3 criteria performance as a function of classification category.

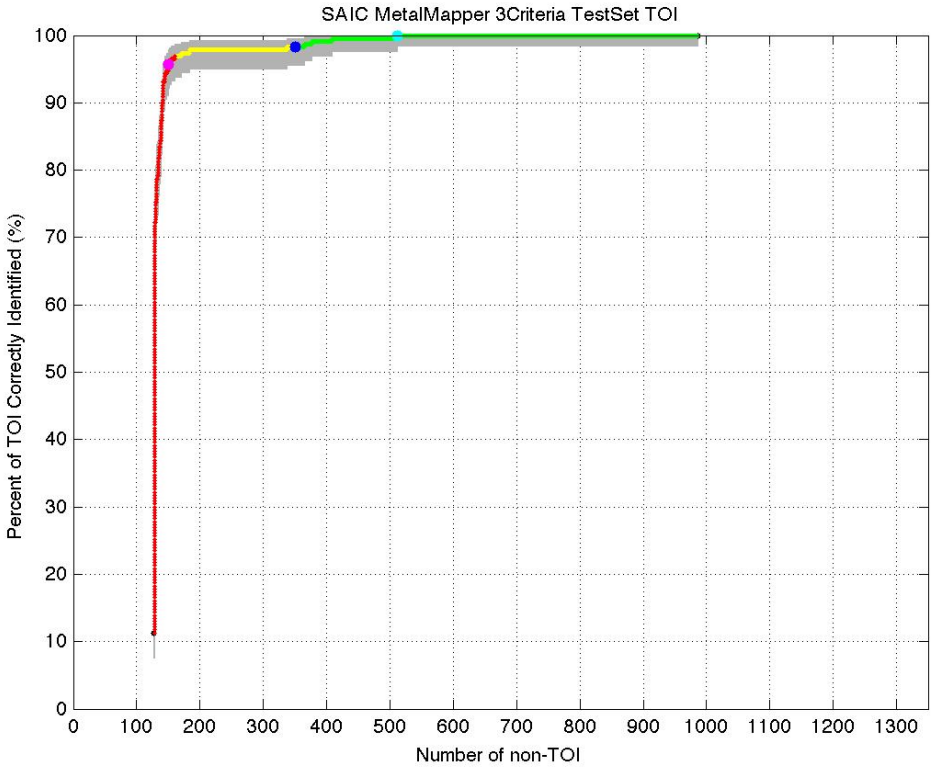


Figure 7-47. Metal Mapper 3 criteria ROC chart.

7.6.2 Characterization Plots

Figure 7-48 shows the difference between the fitted and measured XY locations for all scored category 1, 2 and 3 targets from the Test Set. The mean error for all TOI with an isolated or slightly overlapping signal was 0.12m with a standard deviation of .07m. If non-TOI are added to the population the mean error increases to 0.16m with a standard deviation of 0.18m.

Figure 7-49 shows the difference between the fitted and true depth for all scored category 1, 2 and 3 targets from the Test Set. The mean error for all TOI with an isolated or slightly overlapping signal was 0.03m with a standard deviation of .08m. If non-TOI are added to the population the mean error was also 0.03m with a standard deviation of 0.10m.

In Figure 7-50, we plot the median and 1 standard deviation error bars for our derived β's for all TOI in the Test Set, divided into the four main types. The remarkable consistency shown in these plots validates our direct library matching technique. The 60mm mortars show the most variation overall, due to the presence of a large number of objects without the tail assembly. As with the TEMTADS analysis, the high fidelity data obtained by the Metal Mapper allows the use of shape to characterize the targets instead of only relying on size which was the case for the EM61 sensors.

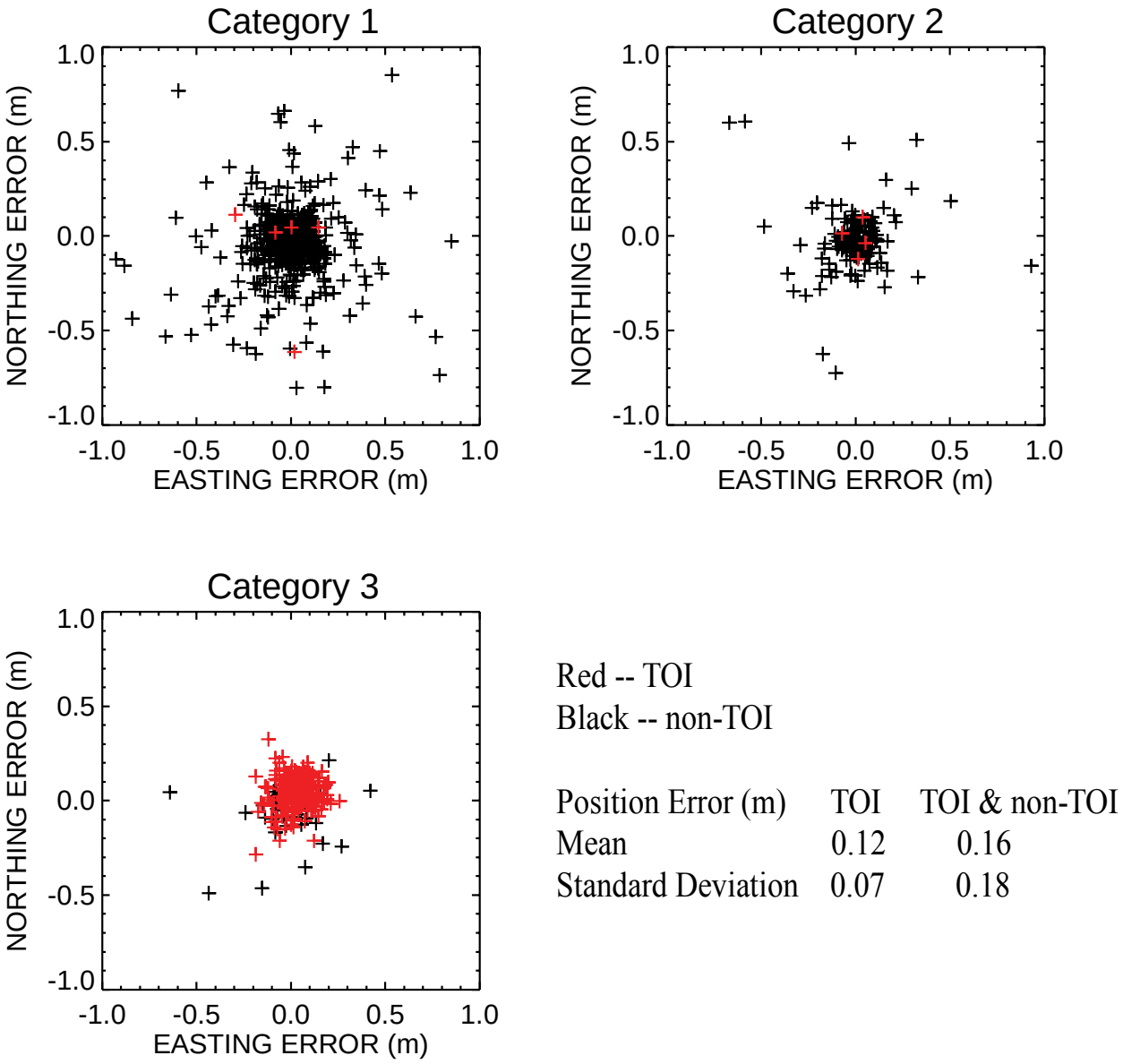


Figure 7-48. Differences between fitted and measured XY locations; Metal Mapper 2 criteria analysis

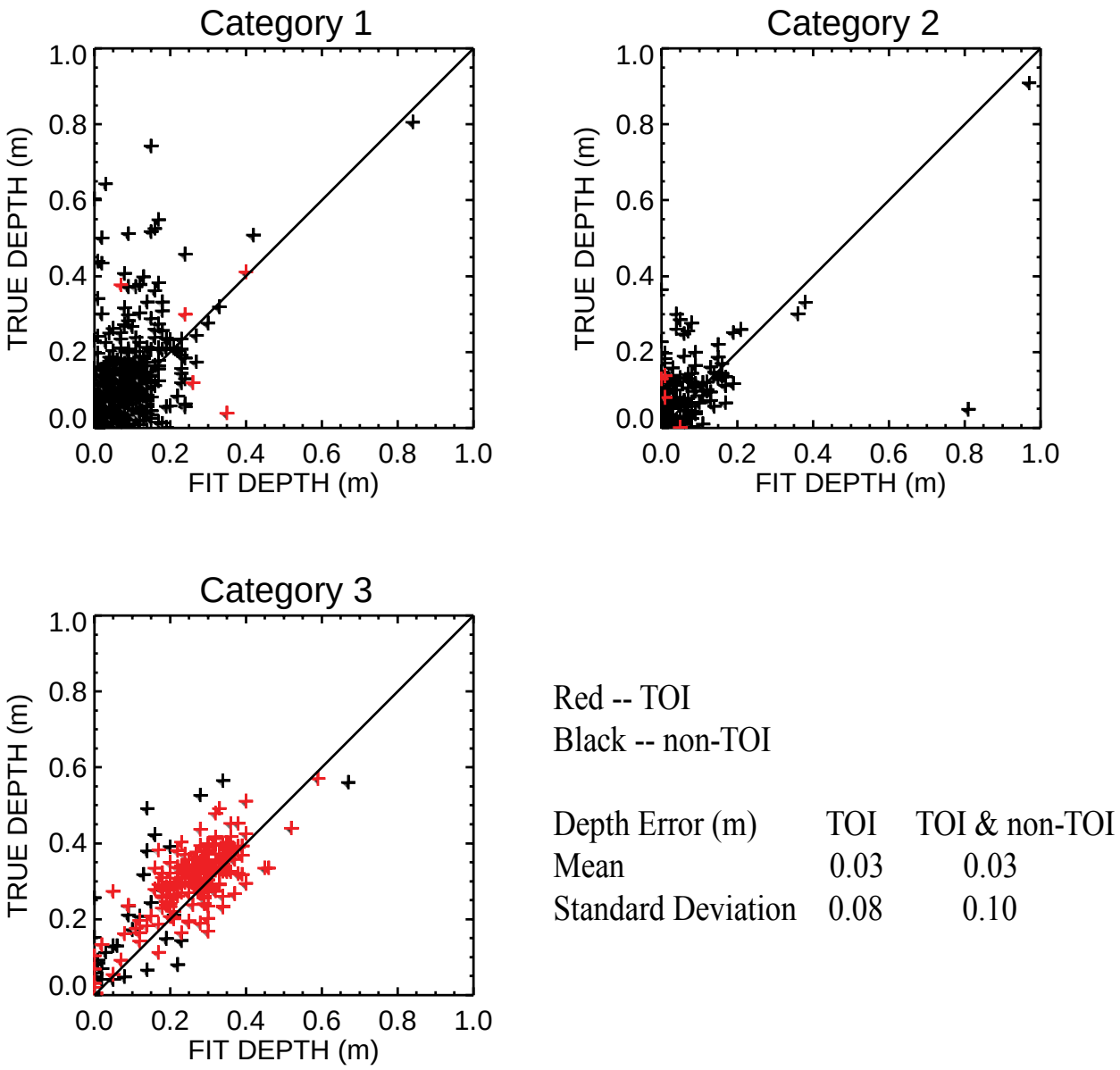


Figure 7-49. Fitted versus measured depth of burial; Metal Mapper 2 criteria analysis

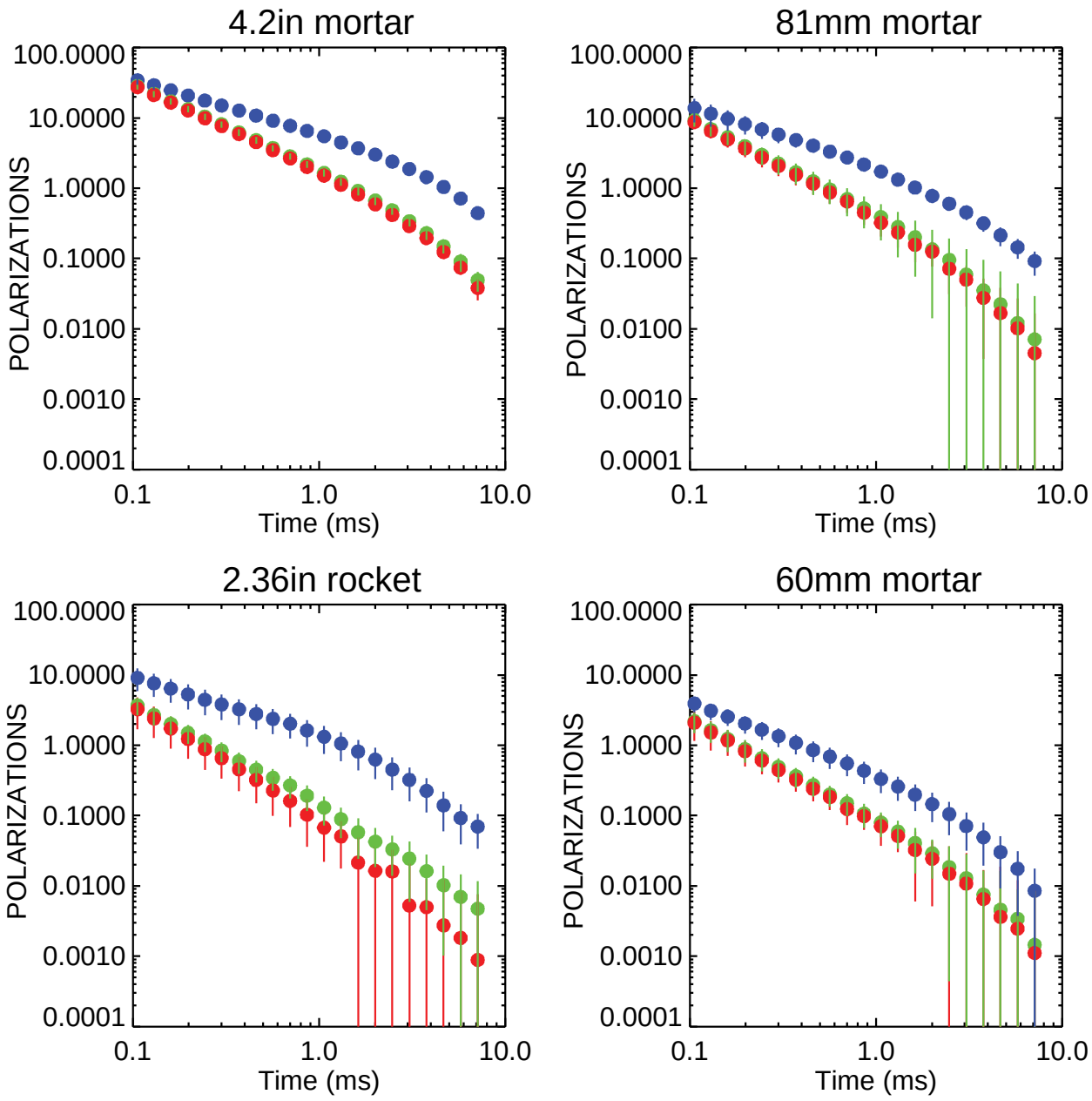


Figure 7-50. Median and ±1σ errors of the four main UXO for the Metal Mapper test set anomalies.

7.6.3 Failure Analyses

In Table 7-14, we give the Metal Mapper and Master IDs of our false negatives for both our 3 and 2 criteria methods. The 2 criteria method contained all 4 false negatives that the 3 criteria method has, along with 1 other.

Table 7-14 Metal Mapper's False Negatives

MM ID	Master ID	3 Criteria/2 Criteria Method
493	314	Both
680	1372	Both
737	1285	2
852	1469	Both
1177	775	Both

Metal Mapper anomaly 1177 (Master ID 775) contains 4 separate objects: a 60mm, 2 tail booms, and a fuze piece, as shown in Figure 7-51 below. As noted previously, our multiple target solver is in its early testing stages, and for this run was set to find a maximum of 3 targets.

Figure 7-51. Photograph of objects found at Anomaly 1177 (Master ID 775).

Anomalies 493 (Master ID 314) and 680 (Master ID 1372) inverted to yield very noisy betas, as shown in Figure 7-52 and Figure 7-53. The problem here appears to be one of low SNR. Since we did not have background measurements for this sensor, we were unable to use these to set a low signal cutoff. Furthermore, the targets in the Training Data had relatively high signal, with the weakest target being stronger than 493, and only 30% weaker than 680. All the Training Data targets inverted well, so these also proved unhelpful in determining a low signal cutoff. Having such a cutoff would allow us to conservatively classify any anomalies below this value as Category 2.

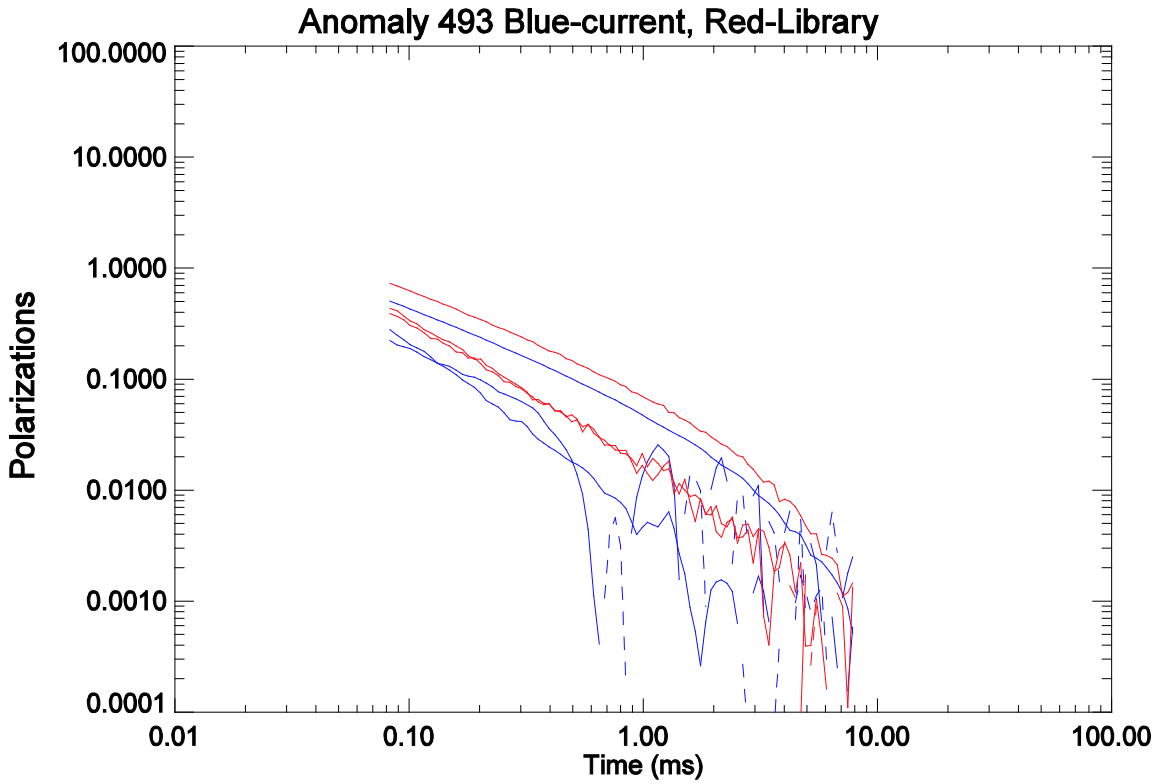


Figure 7-52. Library match to Metal Mapper anomaly 493 a 2.36 inch rocket.

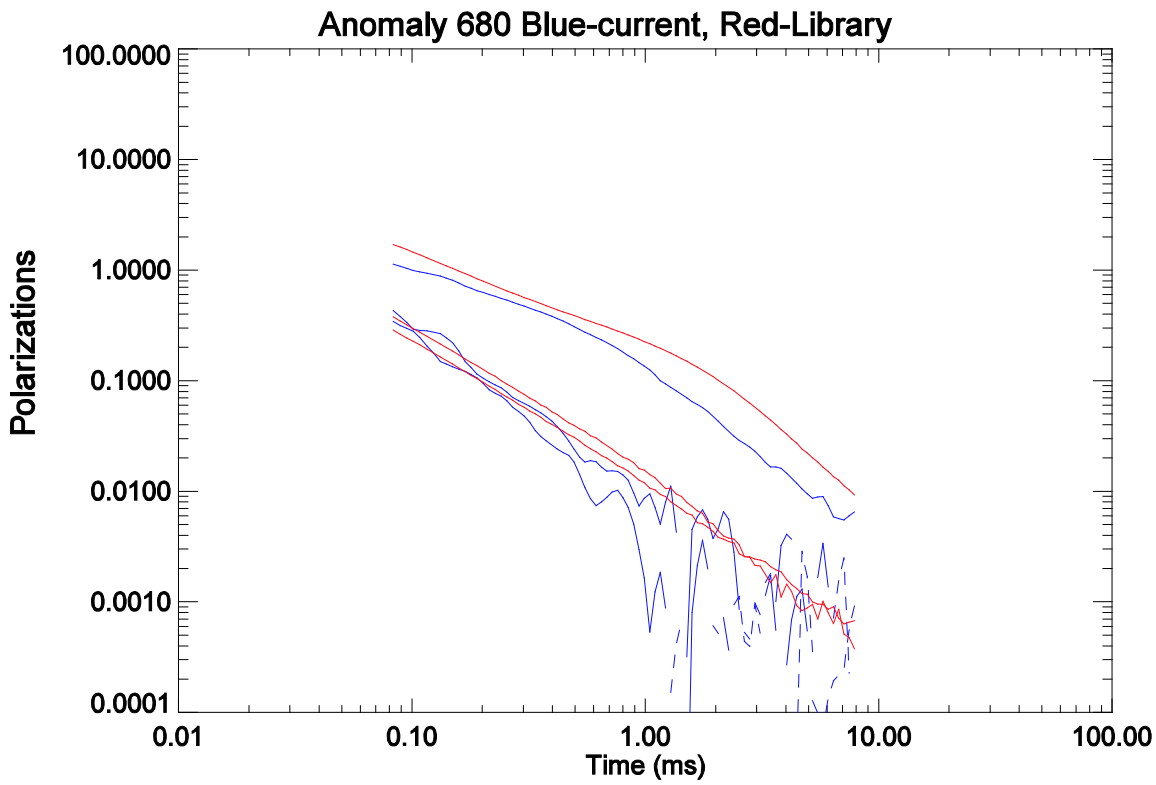


Figure 7-53. Library match to Metal Mapper anomaly 680 a 60mm mortar.

Anomaly 852 (Master ID 1469) is more difficult to explain. As shown in Figure 7-54, although beta 3 is odd in that it falls off too rapidly, none of the betas look particularly noisy. Indeed, the SNR is sufficiently high. There is also no indication of other anomalies present. We suspect our failure for this target to be the result of ordnance variability, as the ground truth lists a fuzed 60mm, and the photo of the target in Figure 7-55 looks rather different than the standard 60mm.

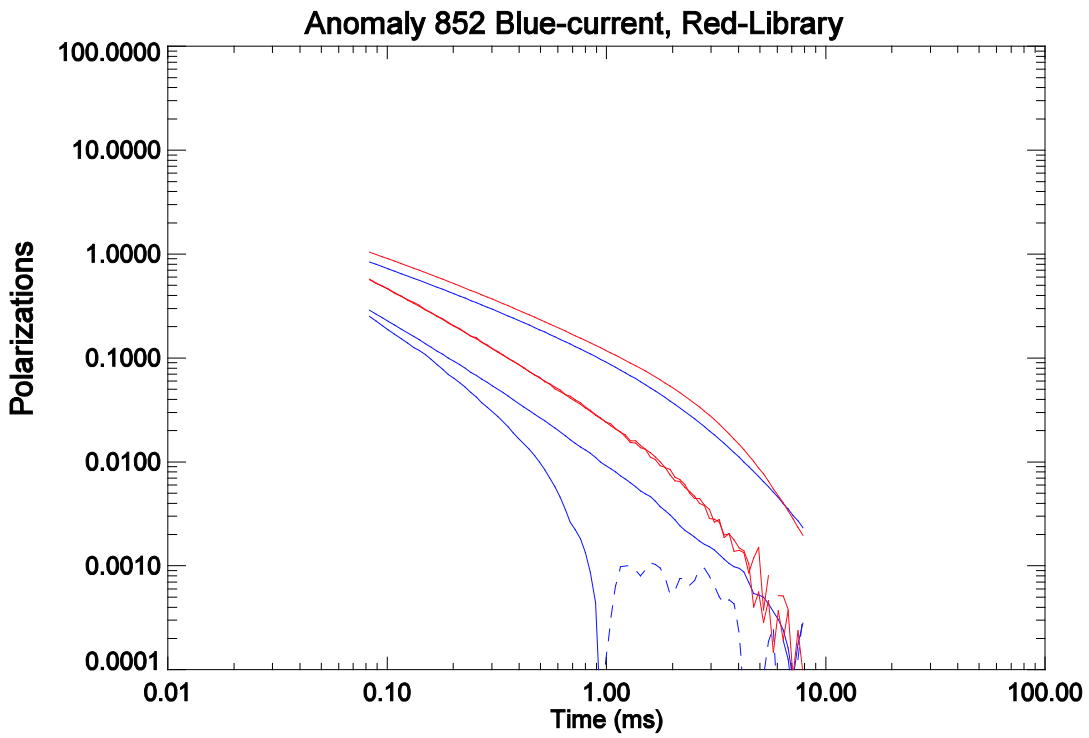


Figure 7-54. Library match to Metal Mapper anomaly 852 a 60mm mortar.

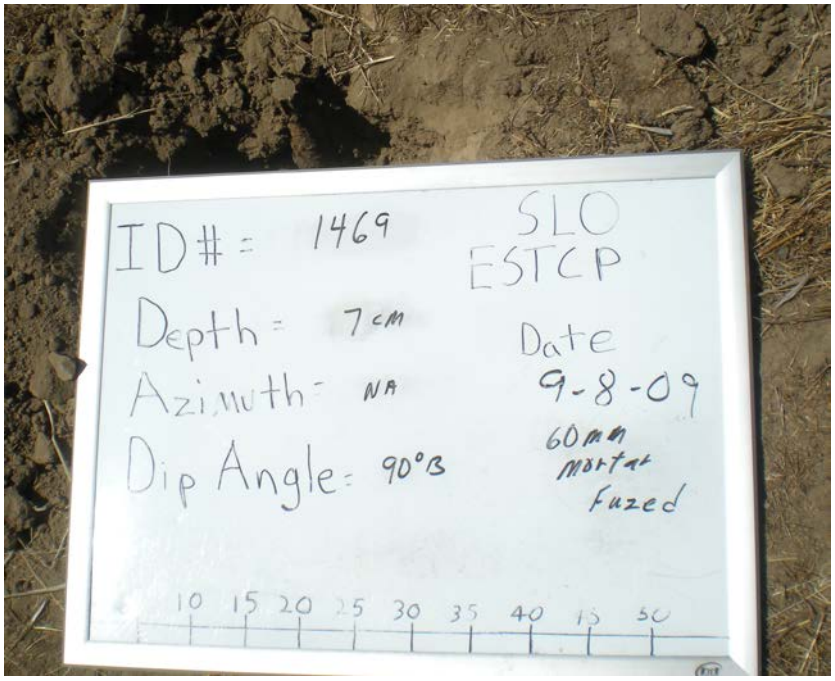


Figure 7-55. Photograph of Metal Mapper anomaly 852, a fuzed 60mm (at the top of the photo) that is different than the 60mm in our library.

The remaining false negative, 737 (Master ID 1285) occurred only for the 2 criteria method. The ground truth lists this target as a 60mm mortar at a depth of 35cm. We show the inverted betas and the best 60mm library match in Figure 7-56. The signal for this target is quite strong; in fact, much too strong to be due to a 35cm-deep 60mm. Our inverted depth for this object places it nearly on the surface. We suspect that there was a piece of surface or near-surface clutter at this anomaly position that was moved subsequent to our data acquisition. We thus were faced with attempting to invert a strong clutter signal with a weak ordnance signal superimposed. Our multi-target solver proved insufficient for the task, finding only the strong surface clutter item.

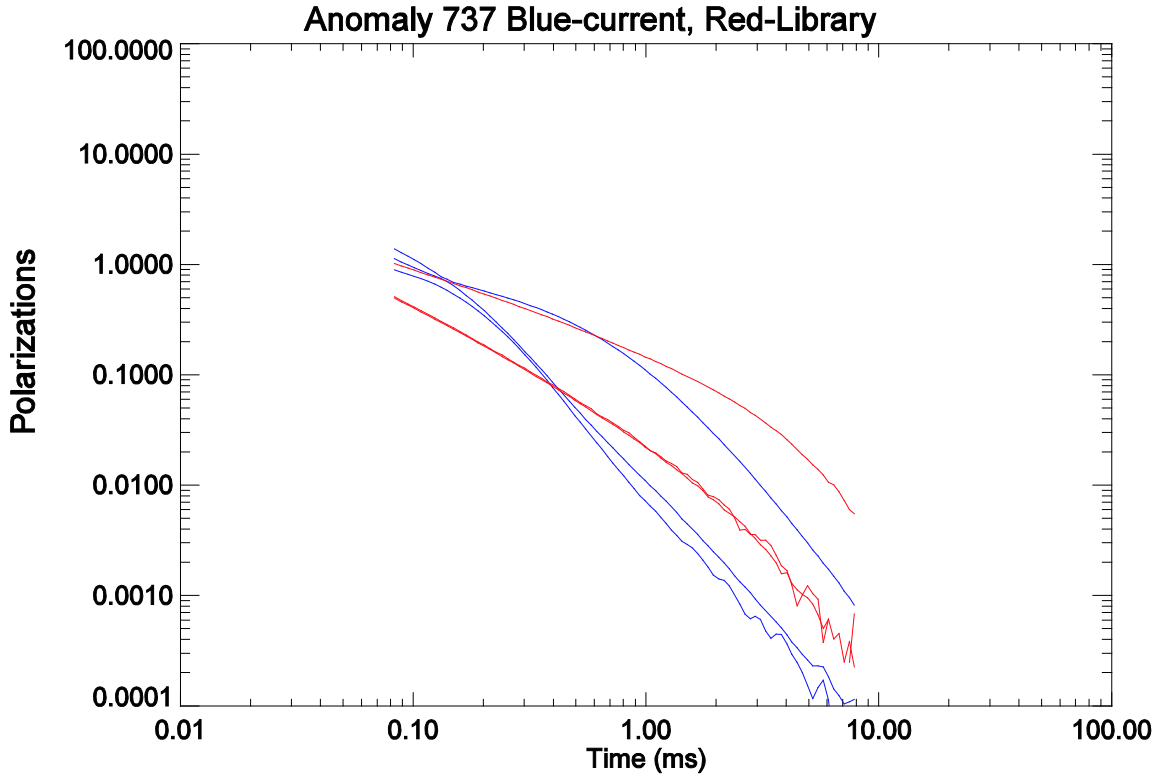


Figure 7-56. Library match to Metal Mapper anomaly 737 a 60mm mortar.

We note that anomaly 737 was classified as Category 2 by the 3 criteria method due to the moderate degree of axial symmetry in the inverted betas as seen in Figure 7-56. The fact that this did not occur for the 2 criteria method was actually due to a flaw in the decision rules logic.

7.7 DISCUSSION

In general, the four data sets collected in dynamic mode using the EM61 sensor produced similar and very good results. Figure 7-57 compares the ROC curves for the SCORR EM61 cart, EM61 MSEMS and EM61 Array data. The curves show minor differences in their shape with the main difference being the user defined threshold. The EM61 Cart and EM61 Array data both rejected similar amounts of clutter (52-55%) while retaining all TOI but the threshold used for the EM61 Array resulted in no false negatives compared to one for the EM61 Cart. The EM61 MSEMS used a threshold that was too aggressive, which resulted in more false negatives. It also had one false negative due to multiple objects towards the end of the ROC curve. All the EM61 data sets used estimated size and decay ratios to classify the anomalies. The individual polarizations were not of sufficient quality to use to calculate accurate shape information.

In contrast, the cued fixed array systems consisting of the TEMTADS and Metal Mapper produced polarizations that were accurate enough to discriminate between TOI and non-TOI on the basis of shape. Figure 7-58 shows the ROC curves for the TEMTADS and the Metal Mapper overlain on the ROC curves for the EM61 sensors. The curves are almost identical for the TEMTADS and Metal Mapper and start off much better than those for the EM61 sensors. There are very few false positives from the beginning of the curve until you reach about 95% of TOI recovered. The remaining 5% of TOI require a large number of excavations and the thresholds defined resulted in 3 and 4 false negatives for the TEMTADS and Metal Mapper, respectively. The increased number of false negatives compared to the EM61 sensors can be explained by a higher confidence in the polarizations for the cued sensors. The increased confidence allows for a more aggressive threshold to be selected but better QC must be performed to ensure the inversion produce trustworthy polarizations.

The false negatives can be divided into two main types. The first is comprised of low SNR targets. These occurred for both the TEMTADS and the Metal Mapper sensors. It was hoped that the 2 criteria method might do better on these cases because the parameter “beta 3” tends to be the first parameter to invert poorly as the SNR decreases, but this was not the case. Now that we have the ground truth for this dataset, improvements for low SNR targets can be handled by undertaking studies to determine a more robust classifier to identify low SNR targets, and therefore safely classify them as Category 2. On the data collection side, tests can be conducted to determine the cost/benefit of increasing the number of stacks for weak targets. More stacking will increase the SNR, but at the cost of longer integration times, and thus fewer targets measured per day.

The second category of false negatives is multiple targets. Improvements here can proceed on two fronts. First, work is already underway to measure and compile a database of overlapping signatures with which we may further train our multi-target solver. In lieu of this, it will be necessary to derive a means of estimating from the data when there is a high likelihood of multiple targets present, and therefore, when we should not trust the results of a single dipole solver and classify the anomaly as Category 2.

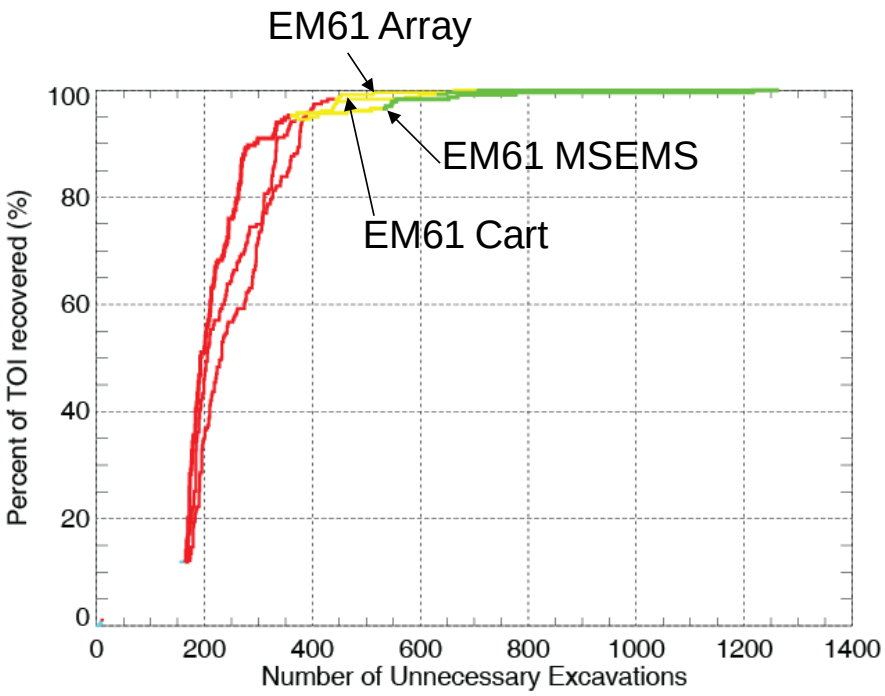


Figure 7-57. Comparison of ROC curves for the three EM61 sensors used in dynamic survey mode.

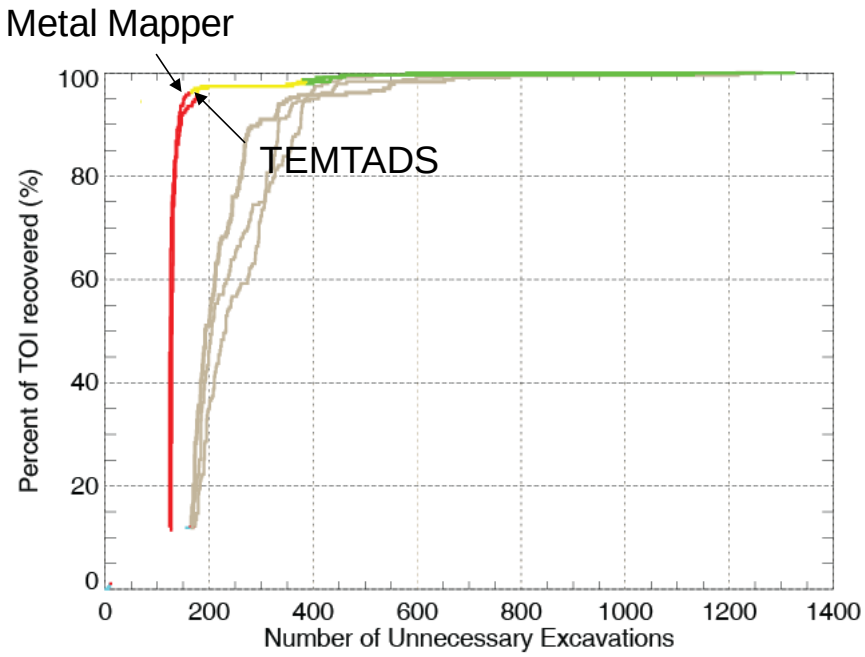


Figure 7-58. Comparison of ROC curves for the two second generation EM sensors used in cued survey mode. The traces in gray are the ROC curves for the EM61 sensors.

The targets analyzed for the TEMTADS were selected from the EM61 Array data by the Program Office and sent to SAIC for analysis. The Metal Mapper targets were selected from the Metal Mapper's dynamic survey by its operators. There were 881 common targets between the two systems. Figure 7-59 show the ROC curves for the TEMTADS and the Metal Mapper for the 881 common targets. The curves are fairly similar with the main difference being that the curve for the Metal Mapper has a steeper rise but the TEMTADS curve reaches 100% of the TOI recovered sooner.

The starting point of the ROC curves presented are offset from the origin. This offset accounts for the real world situation in which all site specific training data required to train the classifiers needed to be dug. As a comparison, we ran a couple of different scenarios that do not require site specific training data. The first scenario which we call the "No on site Learning Necessary" (NOSLN) option assumes knowledge of the TOI munitions present on site but builds the polarization library from a master library of signatures collected at previous sites. The ROC curve for the NOSLN option using TEMTADS data is presented in Figure 7-60. In this figure, there are two NOSLN curves shown and they bifurcate just above the 80% TOI recovery point. The lower of the two curves resulted from the 'phase 1' NOSLN submittal. The higher of the two curves shows the performance for the 'phase 2' NOSLN submittal, which was generated after learning ground truth for all anomalies that we declared high confidence UXO in phase 1. If we compare the NOSLN curve to the original TEMTADS curve we see that they are fairly similar. The NOSLN curve does not rise quite as steep as the TEMTADS curve and the elbow starts a little earlier but those factors are offset by the fact that no on site digging is required to train the classifier and the NOSLN option actually reached the 95% mark earlier. The second scenario assumes no on site knowledge of specific TOI and uses only axial symmetry to classify the targets. Figure 7-61 shows the ROC curve for this option using the TEMTADS data and it is significantly worse than the original TEMTADS curve and the NOSLN option.

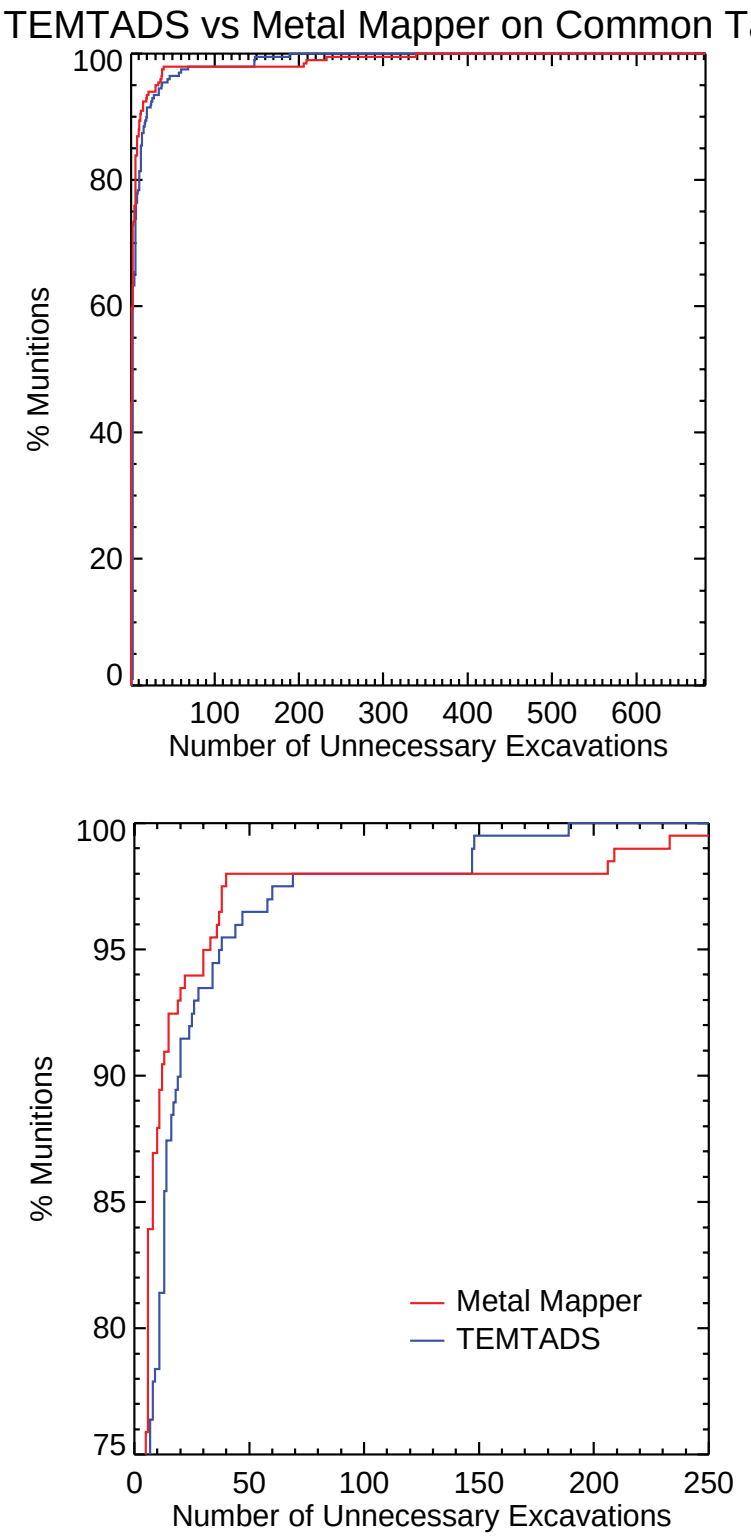


Figure 7-59. ROC curve of TEMTADS and Metal Mapper for common targets. The bottom figure is a zoomed in portion of the upper figure.

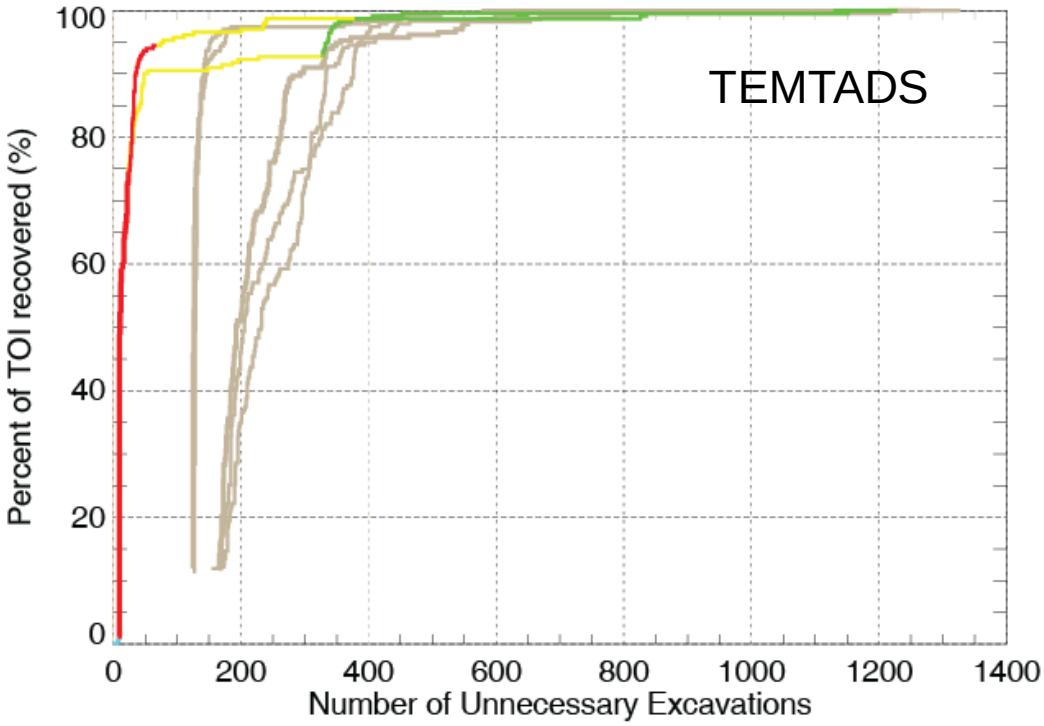


Figure 7-60. The ROC curve of the TEMTADS data assuming knowledge of the TOI munitions type but without on-site training.

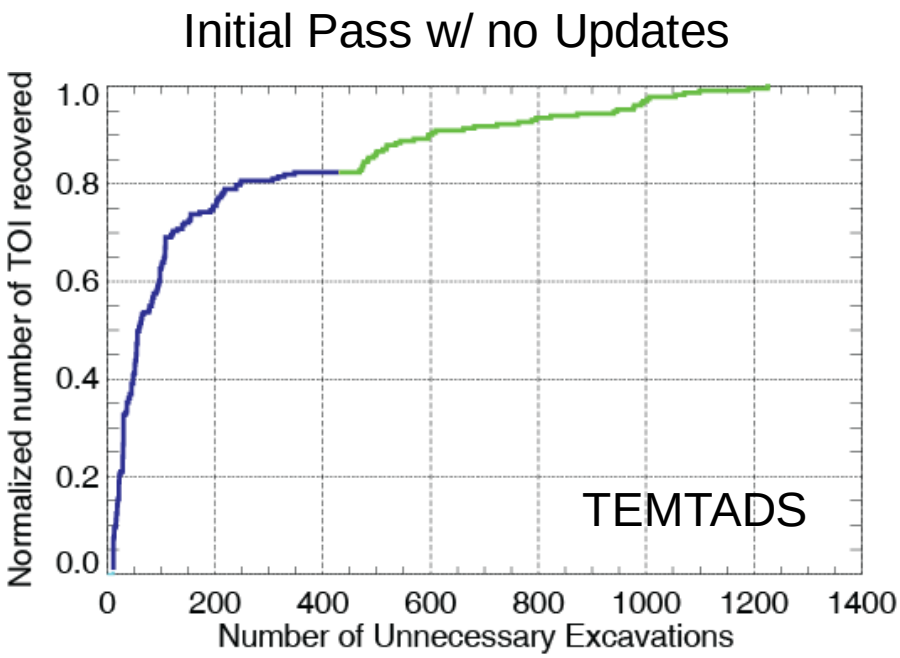


Figure 7-61. The ROC curve of the TEMTADS data assuming no knowledge of specific TOI on site. Classification was solely based on axial symmetry.

8 COST ASSESSMENT

8.1 COST TRACKING

The data analysis performed in this demonstration consists of a number of distinct and sequential operations. The operations include pre-processing, anomaly extraction, characterization (parameter estimation), classification, and documentation. The hours and costs for the individual data sets were tracked throughout the demonstration and are shown in Table 8-1. The reported costs assume an average labor rate of $115 per hour and 1,500 anomalies analyzed.

All the EM61 data sets were analyzed using UX-Analyze. For this reason the methodology and processing speed and therefore the costs were the same for all EM61 data sets. The TEMTADS and Metal Mapper used IDL routines and almost the same process. The TEMTADS inversion routines ran a little slower because the TEMTADS system records more data due to the greater number of coils and time gates. The TEMTADS system also has a 2 square meter footprint compared to the 1 square meter for the Metal Mapper. The larger foot print and the monostatic terms allows the user the ability to visualize the data and select which coils will be used in the inversion. For these reasons the data extraction and parameter estimation costs are more for the TEMTADS system than the Metal Mapper.

The costs associated with “Preprocessing”, “Data Extraction” and “Parameter Estimation” are anomaly specific. It will not make much of a difference to these costs per anomaly if 100 anomalies are analyzed or 10,000 anomalies. The costs will just scale accordingly.

On the other hand the cost associated with “Classifier Training” and “Classification and Construction of the Ranked Anomaly List” are largely independent on the number of anomalies. Once the training data is analyzed and decision rules are developed the difference in time and cost to run 100 anomalies versus 10,000 anomalies through the classifier will be negligible because it is mainly based on computer speed and not operator dependent.

Table 8-1. Cost Summary by each individual data set.

Cost Category	Description	Estimated Costs per Anomaly
Processing Costs (Costs tracked by individual data set)		
Preprocessing	Cost/time required to perform standard cleanup and filtering of the geophysical data.	EM61 Cart - $1 EM61MSEMS - $1 EM61 Array - $1 TEMTADS - $4 Metal Mapper - $4
Data Extraction	Cost/time required to extract data chip encompassing each anomaly	EM61 Cart - $1.5 EM61MSEMS - $1.5 EM61 Array - $1.5 TEMTADS - $1.5 Metal Mapper - $0
Parameter Estimation	Cost/time required to extract parameters for all anomalies and QC results.	EM61 Cart - $3 EM61MSEMS - $3 EM61 Array - $3 TEMTADS - $7 Metal Mapper - $6
Classifier Training	Cost/time required to optimize classifier design and train	EM61 Cart - $1.5 EM61MSEMS - $1.5 EM61 Array - $1.5 TEMTADS - $0.60 Metal Mapper - $0.60
Classification and Construction of the Ranked Anomaly List	Cost/time required to classify anomalies in the test set and construct the ranked anomaly list	EM61 Cart - $0.05 EM61MSEMS - $0.05 EM61 Array - $0.05 TEMTADS - $0.05 Metal Mapper - $0.05
Totals	Total cost to process and classify an anomaly.	EM61 Cart - $7.05 EM61MSEMS - $7.05 EM61 Array - $7.05 TEMTADS - $13.15 Metal Mapper - $10.65

8.2 COST DRIVERS

Data collection: Generally speaking, data collection costs will be greater for classification than detection only. Basically, the EMI analysis process utilizes subtle changes in the anomaly shape. Care must be taken during data collection to not only sample the anomaly fine enough, but also to not introduce noise due to inappropriate collection methods. The costs for data collection vary widely, depending on site conditions such as topography, vegetation, geologic background, known munitions types, as well as weather conditions.

Data Analysis: In general, data analysis costs will be greater for classification than detection only. Data analysis costs are affected by the presence of complex geology, which can make filtering and parameter estimation more complicated. The munitions of interest will also have a great effect on complexity and costs of processing, as will anomaly density. In the case considered here, there were four main TOI and a few areas with high anomaly density. Using the discrimination procedures described in this report we were able classify more than 50% of the non-munitions items correctly with only a handful of false negatives. The number of non-munitions that can be removed with high confidence at another site may be much lower. In addition, the job of the processor in determining the important features and training the classifier may be harder.

Excavation Cost. The costs associated with excavating anomalies vary widely and the goal is to reduce these costs via classification. Safety procedures and nominal burial depth drive remediation costs. When minimal engineering controls are used, costs as low as $45-90 per hole have been reported. When safety procedures are far more elaborate due either to the type of munitions or to their proximity to high value objects, the costs per hole are measured in the hundreds. With regards to burial depth, it is less costly to recover shallow, near-surface items than large deep targets.

8.3 COST BENEFIT

The cost benefit of the classification approach relates to savings realized by not excavating items that are not of interest. The ROC curve in Figure 8-1 shows a three-category classification scheme with a threshold set such that all the items on the right are high confidence non-TOI (Category 1). Although this is an example ROC only, it is very similar in nature to those presented in throughout this report. Note that the anomalies to the right of the threshold were correctly classified as high confidence not munitions. Cost savings can be realized, therefore, if we make use of the classification information and remediate accordingly.

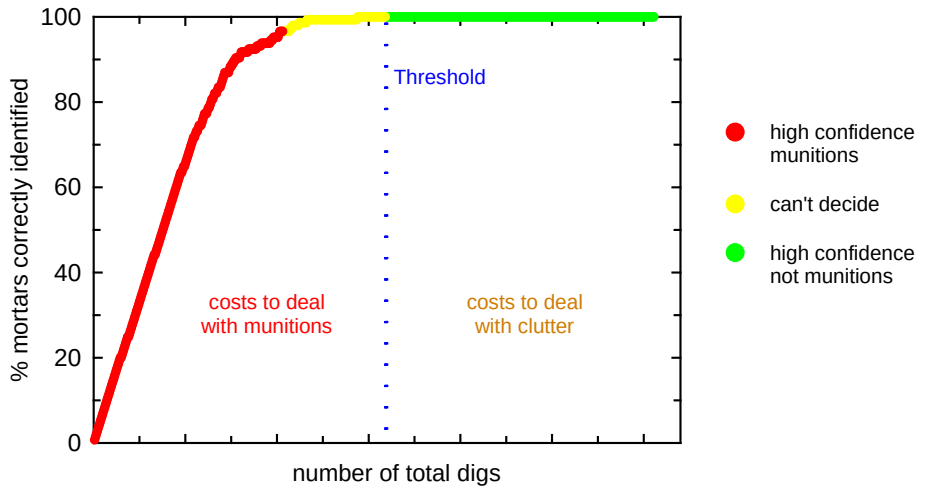


Figure 8-1. Example ROC curve to illustrate cost saving due to proper classification.

Figure 8-2 shows how notional costs accumulate through the process of data collection and processing, digging the munitions, and excavation. In the figure, the detection only (solid black line specifies a lower density data collection for detection only and all anomalies are excavated using intrusive recovery procedures that require trained UXO qualified personnel and safety equipment. The classification 1 (dashed green line) specifies higher density and quality data collection followed by classification processing, and all high-confidence clutter items are left unexcavated. Finally, the classification 2 (dotted green line) specifies higher density and quality data collection followed by classification processing, but a less expensive alternative to the current operational methods of intrusive recovery is used on the anomalies determined to be clutter with high confidence. The classification examples are tied to the different regions of the ROC curve shown in Figure 8-1. There are several important points to note in interpreting this curve: (i) The cumulative cost curves start out on the y-axis at different points. This reflects that the initial costs of higher density data collection and processing for classification are higher than the standard methods. The costs of digging the munitions, which must be borne in all cases, are included here. (ii) The "detection only" curve (solid black line) has a constant slope and ends at the total number of anomalies. All detected anomalies are dug using the same procedures at the same costs. (iii) For both classification examples, all of the items determined to be high confidence munitions or "can't decide" must be dug as though they are munitions. Thus, the two classification examples rise at a slope equal to the detection slope until the threshold is reached on the ROC curve where clutter is identified with high confidence (i.e., the yellow-green transition in Figure 8-1). (iv) In the region where there is high confidence that the remaining anomalies are clutter (green portion of the ROC curve) and it is decided not to dig these anomalies at all, no additional costs are incurred. (v) In the region where there is high confidence that the remaining anomalies are clutter and it is decided to dig these anomalies, but using alternative dig procedures, additional costs are incurred, but the cost of each of these digs is lower so the slope is more gradual. (vi) The break point in cost saving will be determined by the true dollars associated with the data collection, processing and excavation costs – all of which are site specific.

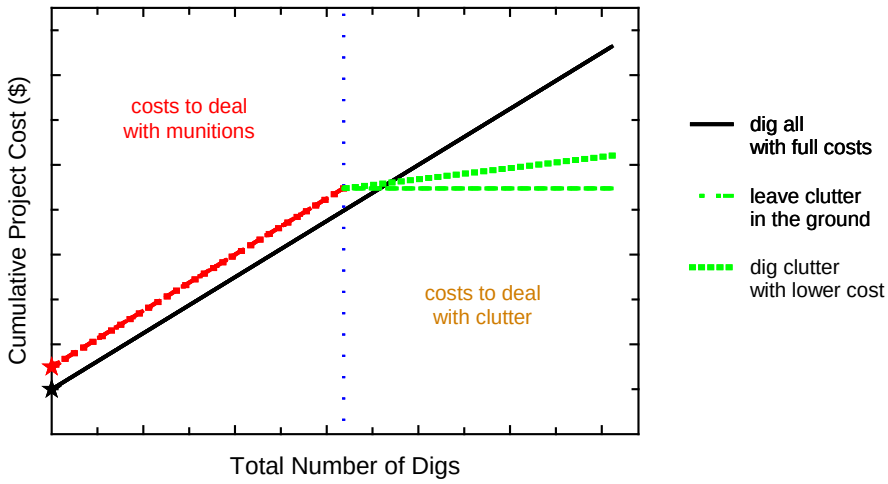


Figure 8-2. Conceptual cost model illustrating the potential savings using the classification methods outlined in this report.

Projecting forward based on experiences from this demonstration, the notional costs for classifying data similar in number and quality to that processed here are approximately $7 for EM61 data and $10 to $13 for the TEMTADS and Metal Mapper. There are additional costs for associated with collecting higher density and quality data for the EM61 based system but the cost savings accrued by digging fewer anomalies is much greater. The TEMTADS and Metal Mapper are cued systems and only required the standard initial detection survey. The added data collections costs only arise from collecting the cued data which is estimated to be about $15 per anomaly. Adding this to the analysis costs results in additional costs of approximately $25-$28 per anomaly for the cued systems. When compared with excavation costs of $50-$100 per hole, the case for advanced classification is still cost effective.

9 IMPLEMENTATION ISSUES

9.1 REGULATORY AND END-USER ISSUES

The ESTCP Program Office established an Advisory Group to facilitate interactions with the regulatory community and potential end-users of this technology. Members of the Advisory Group include representatives of the US EPA, State regulators, Corps of Engineers officials, and representatives from the services. The ESTCP staff worked with the Advisory Group to define goals for this Program and developed Project Quality Objectives. As the analyzed data from the demonstrations become available, the Advisory Group assisted in developing a validation plan.

The discrimination mindset that should be promoted is one that encourages geophysical service providers to deliver a dig list that is prioritized according objectives defined by the site's stakeholders. The decision metrics and thresholds should be quantitative, transparent, and documented. Stakeholder priorities could reasonably relate to size, depth of burial, shape, material type, or munitions type(s). Once the prioritized dig list is delivered, the stakeholders decide the ensuing actions. Stakeholders and/or site managers can realize financial savings by either modifying excavation procedures based upon the probability of being a TOI or by declaring that no further action is required for specific anomalies. The point is that the decision to take action always remains with the stakeholders – as it must, if the discrimination mindset is to be accepted.

10 REFERENCES

1. "Report of the Defense Science Board Task Force on Unexploded Ordnance," December 2003, Office of the Under Secretary of Defense for Acquisition, Technology, and Logistics, Washington, D.C. 20301-3140, http://www.acq.osd.mil/dsb/reports/uxo.pdf.

2. "ESTCP Pilot Program, Classification Approaches in Munitions Response," Nelson, H., Kaye, K., and Andrews, A.

3. "Demonstration Plan for 2008 Classification Study V3" ESTCP Program Office.

4. "MTADS Magnetometer / MkII Demonstration Plan, Former Camp San Luis Obispo," D.A. Steinhurst, submitted to ESTCP Program Office, April, 2009.

5. "MTADS Discrimination Array (TEMTADS) Demonstration Plan Supplement, Former Camp San Luis Obispo," D.A. Steinhurst, submitted to ESTCP Program Office, April, 2009.

6. "Metal Mapper Data Collection Report: Camp San Luis Obispo Discrimination Study" Mark Prouty, submitted to ESTCP Program Office, July 4, 2009.

7. "Use of MSEMS for ESTCP UXO Classification Study at San Luis Obispo, Demonstration Plan" Robert Siegel, submitted to ESTCP Program Office, January, 2009.

8. "Final Site Inspection Report, Former Camp San Luis Obispo, San Luis Obispo, CA," Parsons, Inc., September 2007.

9. "UXO Classification Study: Scoring Memorandum for the former Camp San Luis Obispo. CA," Cazares, S. and Tuley, M., Institute for Defense Analyses, 13 March 2009.

10. "Shape-Based Classification and Discrimination of Subsurface Objects Using Electromagnetic Induction," Thomas H. Bell, Bruce Barrow and Nagi Khadr, Proceedings of the International Geoscience and Remote Sensing Symposium, Seattle, Washington, July 6-10, 1998.

11. "Model-Based Characterization of EM Induction Signatures for UXO/Clutter Discrimination Using the *MTADS* Platform" Bruce Barrow and H. H. Nelson, Proceedings of the UXO Forum 1999, Atlanta, Georgia, May 25-27, 1999.

12. "Source Separation using Sparse-Solution Linear Solvers" Jonathan T. Miller, Dean Keiswetter, Jim Kingdon, Tom Furuya, Bruce Barrow, and Tom Bell. *To be* submitted to SPIE conference on Defense Security and Sensing, Orlando, FL, April, 2010.

Appendix A: Points of Contact

POINT OF CONTACT	ORGANIZATION	Phone Fax e-mail	Role in Project
Dr. Jeff Marqusee	ESTCP Program Office 901 North Stuart Street, Suite 303 Arlington, VA 22203	703-696-2120 (V) 703-696-2114 (F) jeffrey.marqusee@osd.mil	Director, ESTCP
Dr. Anne Andrews	ESTCP Program Office 901 North Stuart Street, Suite 303 Arlington, VA 22203	703-696-3826 (V) 703-696-2114 (F) anne.andrews@osd.mil	Deputy Director, ESTCP
Dr. Herb Nelson	Naval Research Lab Chemistry Division Code 6110 Washington, DC 20375	202-767-3686 (V) 202-404-8119 (F) 202-215-4844 (C) herb.nelson@nrl.navy.mil	Program Manager, MM
Ms. Katherine Kaye	HydroGeoLogic, Inc. 11107 Sunset Hills Road, Suite 400 Reston, VA 20190	410-884-4447 (V) kkaye@hgl.com	Program Manager Assistant, MM
Dr. Shelley Cazares	Institute for Defense Analyses 4850 Mark Center Drive Alexandria, VA 22311	703-845-6792 (V) 703-578-2877 (F) scazares@ida.ord	Performance Assessment
Dr. Dean Keiswetter	SAIC 120 Quade Drive Cary, NC 27513	919-677-1560 (V) 919-678-1508 (F) keiswetterd@saic.com	Project Lead
Dr. Thomas Bell	SAIC 200 12th Street, Suite 1001 Arlington, VA 22202	703-414-3904 (V) 703-416-9306 (F) bellth@saic.com	Quality Assurance Officer
Mr. Tom Furuya	SAIC 120 Quade Drive Cary, NC 27513	919-677-1560 (V) 919-678-1508 (F) furuyat@saic.com	Data Analyst
Dr. Jim Kingdon	SAIC 200 12th Street, Suite 1001 Arlington, VA 22202	703-414-3872 (V) 703-416-9306 (F) kingdonj@saic.com	Data Analyst